

1. コンビナトリアルケミストリ/HTS

私にとり、コンビナトリアルケミストリ/HTS を構造一活性相関手法の一つとして取り上げるには抵抗がある。というのは、コンビナトリアルケミストリ/HTSではリード化合物の発見は行うが、構造一活性相関で最も大事な要因解析を行うものではないし、ドラッグデザインを行うものでもないからである。現に、構造一活性相関とコンビナトリアルケミストリ/HTSとの違いは何ですかと問われる事も多い。そこで本章では本論に入る前にコンビナトリアルケミストリ/HTSと構造一活性相関との差異について一節を設けている。

ここでは、コンビナトリアルケミストリ/HTSがコンピュータを用いる事の多いアプローチであること。また、構造一活性相関の知識がコンビナトリアルケミストリ/HTSの実施に役に立つこと。さらには、構造一活性相関との連携が今後一層重要になるとの観点から、構造一活性相関ではないが関連する重要な技術として取り上げる。現に、コンビナトリアルケミストリではコンピュータが必然的に利用される場面が多々ある。例えば、コンビナトリアルケミストリの基本である“分子多様性 (Molecular Diversity)”に関する基本技術は、第章のパターン認識法を読まれていれば容易に理解できるはずである。また、創出された多数の化合物やHTS (High Throuput Screening)の実施により出てくる大量の薬理データを効率よく管理するにはデータ管理用の専門ソフトウェアが必要となる。また、現時点では実用化されていないが多数の化合物を同時に効率よく合成するための合成デザインにもコンピュータの助けが必要となる。

本書ではコンビナトリアルケミストリにおける化合物ライブラリと、化合物選択の基本となる分子多様性に関する話題と、コンビナトリアルケミストリを考えた時の合成のあり方等に限定して解説する。その他のHTS、および合成に関する詳細な説明は他書を参照されたい。

① コンビナトリアルケミストリ/HTS とは

コンビナトリアルケミストリ/HTSを一言でいうならば、高速なランダムスクリーニング手法である。ランダムスクリーニング自体は最も原始的で最終的なリード化合物発見手法である。しかし、このアプローチは最も非効率的なスクリーニング手法であり、一化合物あたりのスクリーニング単価が沸騰している現在、この手法を取ることは不可能である。唯一例外は、戦時中米国で抗マラリア剤の開発が国家プロジェクトとして採算抜きで行われた程度である。

コンビナトリアルケミストリ/HTSはこのランダムスクリーニングを現代風にリメイクしたものである。このアプローチは、固相合成、コンピュータ、ロボット(合成、スクリーニング)、酵素反応、超微量分析等に関する多方面の技術の集積により実現される。従って、これら総ての技術にボトルネックが

存在し、一面では危うさも持っているが、少なくとも現時点では相乗効果により総ての分野が急速に進歩している状態にある。図 に、コンピュータとロボット及びコンビナトリアルケミストリ/HTS との関係を簡単に示す。

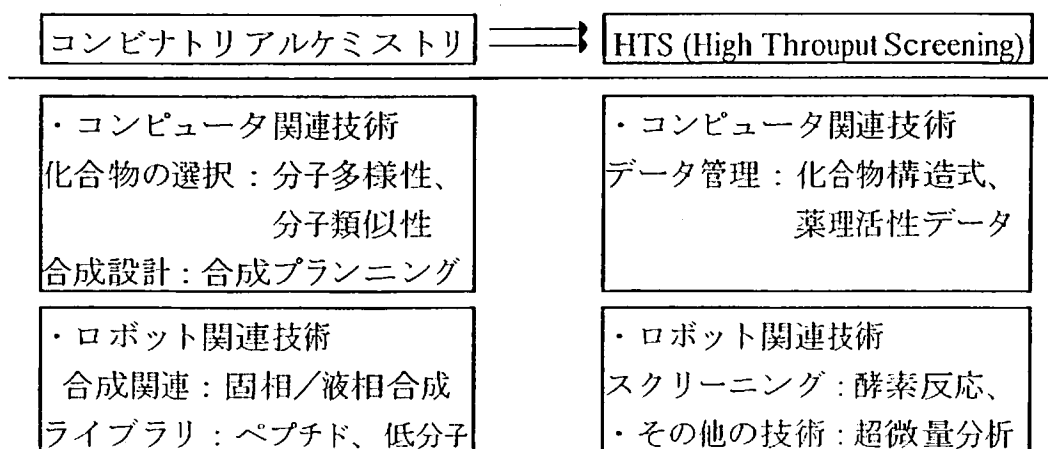


図 . コンピュータ、ロボットとコンビナトリアルケミストリ/HTSの業務

② コンビナトリアルケミストリ/HTS と構造-活性相関の違い

・役割の違い

両手法の比較はランダムスクリーニングと構造-活性相関の違いと同値である。古来よりリード化合物発見の究極手法としてランダムスクリーニングが存在した。これは、ランダムスクリーニングが実験による薬理活性の評価を基本とし、その実験結果は絶対的な重みを持つためである。しかし、スクリーニングコストが高くなると、無作為に選択された化合物を対象とする本手法はコスト的に実用に耐え得られなくなる。これがランダムスクリーニングの限界である。

スクリーニングコストを下げられない以上、取り得る手段はスクリーニング対象となる化合物のヒット率を高めること以外に存在しない。このヒット率を高めることが、総ての構造-活性相関手法の主要な目的の一つである。しかし、リード化合物探索のヒット率は構造-活性相関により向上出来るが、このヒット率はあくまでも推論にしかすぎない。手続きの複雑さと、推論であるがための説得力の弱さが構造-活性相関の弱点である。

- | |
|--|
| <ul style="list-style-type: none"> ・ランダムスクリーニング：実験主体の実証主義に基づくアプローチ 限界：ヒット率が低く、非能率的、高コスト体質 ・構造-活性相関：解析主体の理論による化合物のヒット率向上が目的 限界：ヒット率は上昇、あくまでも推論である、低コスト化可能 |
|--|

ランダムスクリーニングの限界項目中、非能率的、と高コスト体質を最新の

技術で改善したものがコンビナトリアルケミストリ/HTSである。もう一つの限界であるヒット率に関してはこれら二項目の改善で、ある程度カバー出来る。ランダムスクリーニングが実証主義に基づき、これ以上に確実なスクリーニング手法が無い現状では、その限界が除かれたならば堰を切ったように多くの企業が採用するのは当然のことである。

・ アプローチの違い

コンビナトリアルケミストリ/HTS はリード化合物の発見業務に特化したアプローチである。一方、構造-活性相関はリード化合物を発見するという目的も大きな課題ではあるが、これ以外にも薬理活性を向上させる、毒性や副作用を除く、要因解析により薬理活性メカニズムを明確にする等の多様な目的を持つ。ここでの二つのアプローチの比較は、リード化合物発見業務に限定して述べる。つまり、コンビナトリアルケミストリ/HTS と構造-活性相関のリード化合物発見に対するアプローチの基本的な差異について、簡単な比喩により説明する。

いま、銃を用いてターゲットを撃つ場合を想定する。ターゲットはリード化合物であり、射手の周辺空間に散在するものとする。銃の種類や撃つための様々なコンディションがアプローチの差異を示すものとする。この時、コンビナトリアルケミストリ/HTS はマシンガンやショットガンを用いて周辺の空間に狙いを定めることなく均等に鉄や鉛の弾を高速、かつ多数撃ち込む手法に例えられる。一方、構造-活性相関は周辺空間中にゴーストを浮かび上がらせ、このゴーストめがけて金製の弾を用い、ライフル銃を用いて一発必中の心構えで少数の弾を狙い撃ちするアプローチである。この関係を第 図にしめす。

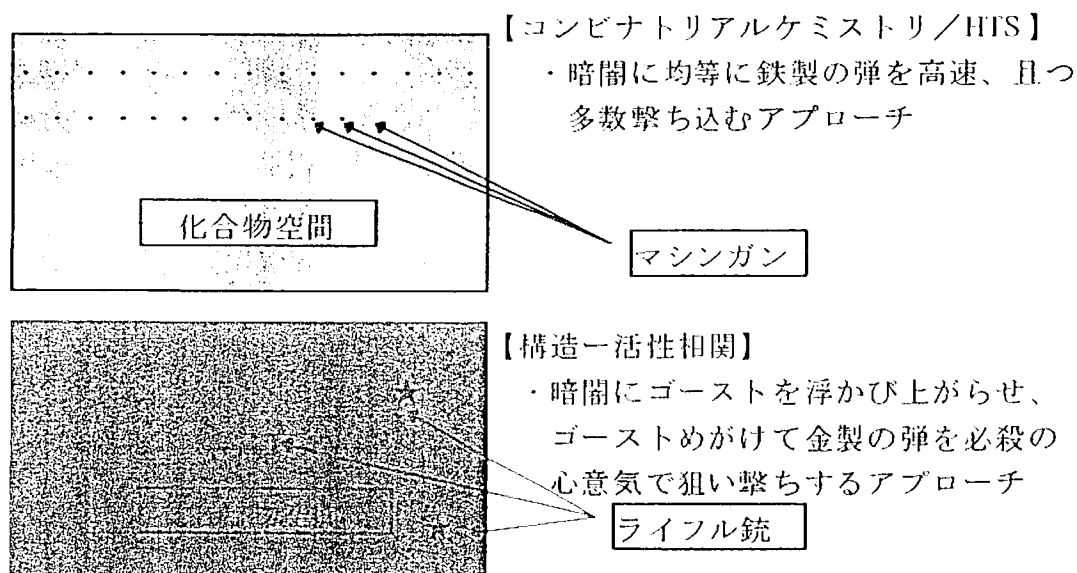


図 . コンビナトリアルケミストリ/HTS と構造-活性相関の違い (比喩)

この比喻について説明すると、周辺の空間はリード化合物を含む化合物空間でどこにリード化合物があるかは不明である。銃の種類はスクリーニングの早さを象徴する。また、利用される弾の原料（鉄／鉛と金）はスクリーニングにかかる一化合物あたりの費用を象徴する。ゴーストは構造－活性相関解析により、リード候補化合物としてデザイン／選択された化合物を意味する。

コンビナトリアルケミストリ／HTSの本質はスクリーニング速度と数にある。このスクリーニングが安価で効率よく、かつ高速に行われるならば先にも述べたようにこれに変わりうるリード化合物発見手段は無い。ちなみに、現時点ではロボットや関連装置の装備度の程度により差があるが、大体数千～数万化合物／週から数十万～数百万化合物／年の処理能力を持っている。これを先の比喻に例えれば、これだけの数の弾を撃てるマシンガンを手にしたことになる。

一方、構造－活性相関の目的は暗闇（リード化合物が潜む化合物空間）に目標となるリード化合物のゴーストを如何にしてゴーストから現実のものとするかであり、この時前提とする銃はライフル銃のような精度重視のマシンで、しかも高価な弾を用いることになる。

- ・ リード化合物発見に対する業務内容の差（相補、および協調関係）

コンビナトリアルケミストリ／HTSは、リード化合物をスポット的に発見するという単純な発見業務しか行わない。しかも、一つのリード化合物の発見はその次のリード化合物発見と結びつかない全く独立した作業となる。また、この他の様々な業務、例えばより活性の高い化合物への誘導、毒性／副作用の無い化合物への誘導、関連化合物の創出、あるいはリード化合物発見後の特許対策等のドラッグデザインに関する業務は一切行わない。構造－活性相関は逆にこのような要因解析やデザイン業務に優れたアプローチである。従って、新薬を開発するという業務全体を考えた場合、コンビナトリアルケミストリ／HTSと構造－活性相関は対立関係にあるのではなく、むしろ相補、および協調関係にあると言える。

更には化合物ライブラリの項でも述べるが、構造－活性相関で築き上げられた技術やノウハウを利用することでコンビナトリアルケミストリ／HTS自体の業務の効率向上が実現出来る。例えば、化合物ライブラリの構築に構造－活性相関の技術を応用することでリード化合物リッチなライブラリを導くことが出来る。つまり、・項で述べたがコンビナトリアルケミストリ／HTSの大きな限界であるヒット率が低いという問題が構造－活性相関技術の導入により解決可能となる。このように、コンビナトリアルケミストリ／HTS自体の高効率化にも構造－活性相関の技術は貢献する。

また、コンビナトリアルケミストリ／HTSによりリード化合物が多数発見さ

れるようになれば、今以上に構造-活性相関の重要性は増大する。勿論、この場合は構造-活性相関自体もコンビナトリアルケミストリ/HTS との協調作業に特化したものへと変化するか、新たな手法が開発されるものとする。このコンビナトリアルケミストリ、HTS と構造-活性相関との関係を新薬開発業務の流れに合わせてまとめたものを図に示す。

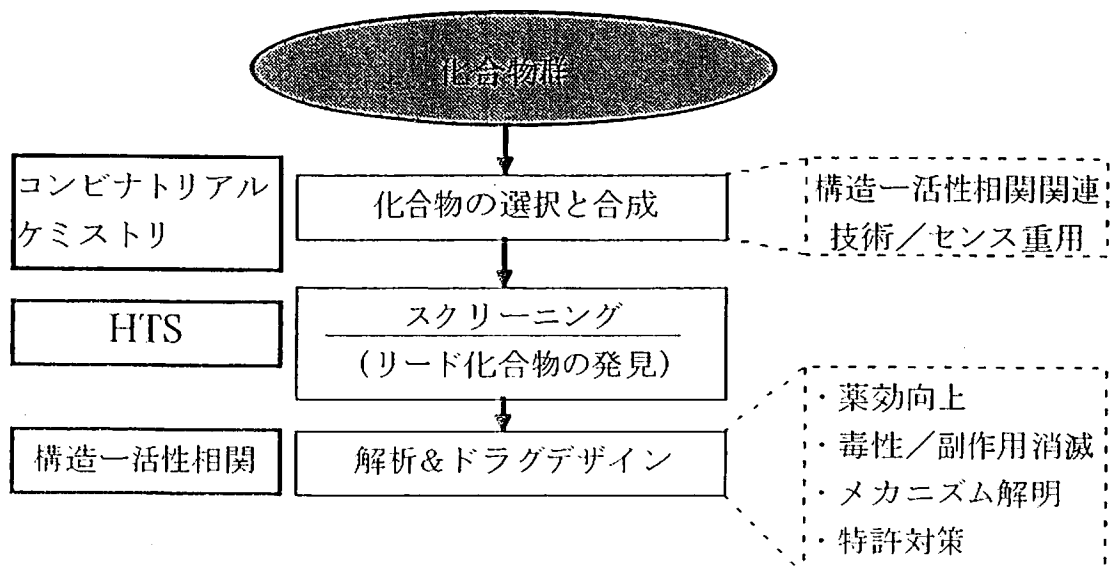


図 業務の流れにおけるコンビナトリアルケミストリ、HTS および構造-活性相関との関係

2. コンビナトリアルケミストリ/HTSの要素技術

コンビナトリアルケミストリ/HTS は多種多様の技術の集積により一つの形を形成している。コンビナトリアルケミストリ/HTS 全般についての解説は既に良書が出版されているのでそちらを参照されたい^{*1)}。

*1: コンビナトリアルケミストリー研究会、'コンビナトリアルケミストリー (入門から応用まで)'、コンビナトリアルケミストリー研究会編、化学同人、東京、1997。

この業務はこの言葉通りにコンビナトリアルケミストリ部分と HTS (High Throuput Screening)の二つの部分に大きく分けて議論することが出来る。この二つの部分は内容により、更に細かく分類される。以下に、コンビナトリアルケミストリ/HTS を構成する技術や項目について簡単に列挙する。

- ① コンビナトリアルケミストリ (スクリーニング対象化合物の調整)
- ・化合物ライブラリデザイン
 - ・化合物の選択技術: 分子多様性、パターン認識、化合物データベース、他
 - ・ライブラリの種類: ペプチド/ペプトイドライブラリ、低分子ライブラリ、試薬ライブラリ、社内化合物ライブラリ、他

- ・ライブラリの内容：分子多様性ライブラリ、リード候補化合物ライブラリ
- ・化合物合成
- ・合成デザイン：合成プランニング、反応データベース、他
- ・合成技術：固相合成、液相合成
 スプリット合成、パラレル合成
- ・合成ロボット：ペプチド／ペプトイド合成ロボット、固相合成ロボット、
 液相合成ロボット
- ・化合物の管理
- ・化合物データベース、合成や HTS ロボットとの連動

② HTS

- ・ロボット／酵素反応関連
- ・アッセイロボット（96 穴、384 穴）
- ・酵素の発見／決定／調達、市販酵素、人工酵素の合成
- ・検出関連技術
- ・分光分析：蛍光法、化学発光／生物発光、UV、NMR、
 RI (Radioisotope : SPA (Scintillation Proximity Assay))
- ・分取／検知：GC、HPLC、Mass、他
- ・スクリーニングデータの管理
- ・薬理活性データベース、合成や HTS（アッセイ）ロボットとの連動

この内容を見ればコンビナトリアルケミストリ／HTS が多種多様の技術で結集された総合技術の結晶である事がわかれる。しかも、ここで取り上げられた多くの技術は互いに殆ど独立して展開されてきた技術である。コンビナトリアルケミストリ／HTS は、まさに化学分野の自動車とも言えよう。これらの技術の個々の内容についてはそれぞれの分野での専門書を参考にされたい。本書ではコンビナトリアルケミストリの部分について言及するが、合成部分を除き、特にスクリーニング対象となる化合物群（化合物ライブラリ）の観点から、その化合物調整に関する基本技術となる分子多様性 (Molecular Diversity) を中心にまとめる。

3. 化合物ライブラリ

3.1 化合物ライブラリの重要性と問題点

コンビナトリアルケミストリの最終目的は、スクリーニング対象となる化合物の選択と調整（合成）である。この場合、大きな問題となる事項は二つある。一つは無量大という数との戦いである。事実上無量大に存在する膨大な化合物群総てをスクリーニングする事は不可能である。現時点での最高環境で達成さ

れる数百万化合物/年レベルで作業を行っても、無限大から見れば宇宙の星屑の一つにもならない。残るもう一つの問題はコンビナトリアルケミストリ/HTS は実証主義に基づくアプローチであることに起因する。つまり、実証主義であるが故にリード化合物発見においてこれに代わりうる強力なアプローチは存在しないが、スクリーニング対象の化合物は必ず合成する事が必要となる。紙に書かれた構造式だけでは実験が出来ないのである。

以上の二点を考えるならばコンビナトリアルケミストリ、特に化合物ライブラリ構築時における目標が明確となる。つまり、より少ない化合物のパターンを用いて、可能な限り多数の化合物群をスクリーニングしたのと同じ効果を得ること（分子多様性の利用）。且つ、ライブラリとして選択された化合物群が、理想をいえば総てリード化合物であること（構造-活性相関技術の適用）である。いかに困難な合成をこなし、多数の化合物群のスクリーニングを行ったとしても、もともとスクリーニング対象の化合物（以下パターンと呼ぶ）群中にリード化合物が無ければ全作業が無駄になる。従って、この化合物選択過程がコンビナトリアルケミストリ/HTS の成否を握る最重要過程となる。

この、化合物の化合物ライブラリの構築過程で様々な問題をクリアすることが求められる。このクリアすべき問題として大まかに、①サンプリングの問題、②リード化合物含有率の問題、③合成の問題、の三種類がある。例えば、サンプリングの問題では探索効率の向上と、探索範囲を大きくするという観点から、化合物のパターン群は重複がなく、かつ可能なかぎり偏りのないことが求められる。また、最重要問題として、化合物ライブラリ自体の品質（リード化合物の含有率）を高めることが必要である。海岸の砂浜から一個のダイヤモンドを探すのと、砂場の中から探すのでは発見効率は大幅に変化する。これが②のリード化合物含有率の問題である。この他、実際に合成することを考えて化合物群を選択することも考慮しなければならない。例え化合物がリード化合物であったとしても、合成できなければ意味がない。①のサンプリング問題に関しては、分子多様性とパターン認識に関する技術が重要となる。②は構造-活性相関技術の応用により有る程度実現可能である。③の問題は合成のスケジューリングシステムに関する問題と、反応データベースにわけられる。

- | | | | | |
|---------------|---|--------------|----------|--------|
| ①サンプリング | ・ | 分子多様性、 | 分子類似性、 | パターン認識 |
| ② Lead 化合物含有率 | ・ | 構造-活性相関技術の利用 | | |
| ③合成 | ・ | 合成スケジューリング、 | 反応データベース | |

本書では、①と②の次項について簡単にまとめる。③の問題は、現在固相合成のデータベースが実現されている程度で、コンビナトリアルケミストリ/HTS を目指した合成スケジューリングシステムは今後急速に立ち上がってくるものと予想する。これらの合成に関する話題については他書に譲る。

3.2 化合物ライブラリの種類と内容

化合物ライブラリはライブラリ中のパターンの種類、起源、入手経路および内容（サンプリング、リード化合物の含有率）により以下のように分類される。

表 様々な基準による化合物ライブラリの分類

分類基準	化合物ライブラリ
種類	ペプチド・ペプトイド、 一般低分子化合物
起源	試薬、 過去の合成化合物、 新規合成化合物
入手経路	市販化合物、文献既知化合物、自社合成化合物
内容	分子多様性考慮、 構造-活性相関考慮

現時点（1997年）で実用化され、最も普及しているライブラリは、自動合成ロボットを用いたペプチドおよびペプトイドライブラリである。ペプチドの自動合成ロボットは比較的古くから実用化されており、このロボットがコンビナトリアルケミストリ/HTSに適用され、ペプトイドへと拡大適用されたのは当然の事である。しかし、薬と言う立場から考えるならばペプチドやペプトイドは代謝の問題等から薬の真打ちと考えるににくく、真の真打ちは一般低分子にある。この一般低分子ライブラリとして簡単に考えられるものとしては市販の試薬を用いる事である。この試薬ライブラリとして、予め分子多様性の評価を行って試薬化合物群から構造的な重複がないように選択された“試薬”ライブラリが販売されている。また企業において最初にスクリーニングすべき化合物ライブラリは、他社に対して競争力のある自社合成化合物群であることは当然である。このスクリーニングが完了したならば、先の試薬ライブラリや、その他の化合物群に対する調達や合成が必要となる。

コンビナトリアルケミストリ/HTSの普及に伴い、従来は大部分が見捨てられていた大学や公共機関等で合成された化合物群の見直しが始まっているようである。また、固相合成自体も急速に裾野が拡大し、最近の学会発表でも固相合成関連の発表が急増している。固相合成に基づく系統たてた化合物合成技術の確立はこれから展開される分野である。

3.3 化合物ライブラリの内容からみた二種類のライブラリ

化合物ライブラリを構築する時の目標内容に従ってみると、化合物ライブラリは大きく二種類に分類される。一つは分子多様性の考慮によるサンプリング重視のライブラリであり、一つは構造-活性相関を考慮したリード化合物の含有率を重視したライブラリ（バイアスドライブラリ(biased library)と言うが本著では“リード候補化合物ライブラリ”として統一する）である。

サンプリング重視のライブラリは化合物の情報（構造）重複を許さないこと

が目的であり、分子多様性のみに基づき選択された化合物群で、目標薬理活性を特定しない。この例として一般の試薬ライブラリ等が該当する。

これに対し、“リード候補化合物ライブラリ”は分子多様性による化合物構造の重複回避と同時にリード候補化合物の存在率も考慮するもので、このような特性を持つ化合物ライブラリの構築には構造-活性相関（特にパターン認識法）に関する技術が非常に重要である。

3.4 分子多様性 (Molecular Diversity) に基づく化合物選択の問題と選択手順

3.4.1 化合物選択 (サンプリング) の重要性和 “組み合わせの爆発” の問題

ここでは化合物のサンプリングに関係する問題を考える。最初の問題はコンピュータ分野では一般的な“組み合わせの爆発”に関するものである。化合物の構造式は無限に創出可能であり、いかにコンビナトリアルケミストリ/HTSの処理速度が早いものであろうと、その処理化合物数は微々たるものにすぎない。この意味で、化合物の選択技術（ライブラリの構築のためのサンプリング技術）の良否はコンビナトリアルケミストリ/HTSの実験を効率良く実施するうえで非常に重要な役割を担うと同時に、その成否をも握る重要な過程である。

◎ “組み合わせの爆発 (Explosion of Combination)” 問題

化合物ライブラリの構築で一般的な、基本骨格と部分構造との組み合わせで化合物を創出する時の克服しがたい大きな問題として“組み合わせの爆発”に関する問題がある。この問題はコンピュータを扱っているとしばしばぶつかる深刻な問題である。

例えば、五カ所に置換基を持つ基本骨格がある時、その置換基を20通りに変換可能ならば、理論上合成出来る化合物数は $20^5 = 320$ 万（ペプチドならばペプタペプチドの合成に該当）となる。この数は、現時点における最新/最強のアッセイロボットを使ってもスクリーニングだけで一年かかることを意味する。置換位置が六カ所となると二十年必要で、わずか七カ所で四世紀もの時間が必要となる。さらに基本構造式自体の変化性を考慮するならば宇宙スケールでの時間が必要となる。

このように化合物の自由度を考えるならば、コンビナトリアルケミストリ/HTSで扱える化合物数は微々たるものである。この意味でも、化合物の選択技術の重要性は非常に高く、リード化合物発見の成否を握るものとなる。以下、化合物選択の基本技術である分子多様性に基づいた化合物選択と、構造-活性相関の技術を適用する事でリード候補化合物の存在比を高め、コンビナトリアルケミストリ/HTS データベースの弱点とされたリード化合物のヒット率を高める技術について述べる。

3.4.2 分子多様性に基づく化合物の選択

分子多様性は化合物選択技術の基本となるもので、化合物ライブラリの構築に利用される。この言葉の類義語として分子類似（相似）性（Molecular Similarity）があるが、化合物選択の観点からは同義語と見なして構わない。本書では特にことわる必要がある時以外は分子多様性の言葉を用いる。

分子多様性とはいわゆる化合物の構造的な差異のバラツキを意味し、化合物の類似度及び非類似度に関する指標で表現される。コンピュータを用いて分子多様性を評価する限り、研究者が構造式を見て分類するわけには行かない。コンピュータで扱えるように構造式情報を変換し、変換された情報を用いて分子多様性を評価する為の指標を計算することが必要である。この化合物の類似／非類似に関する一連の技術の基本が化学パターン認識である。化学パターン認識研究の分野では化合物の類似／非類似の問題はかなり古くから展開されてきた典型的な問題の一つである。従って、化学パターン認識を理解していればさほどの抵抗感なく分子多様性を議論する事が出来る。

◎分子多様性による化合物ライブラリ構築手順

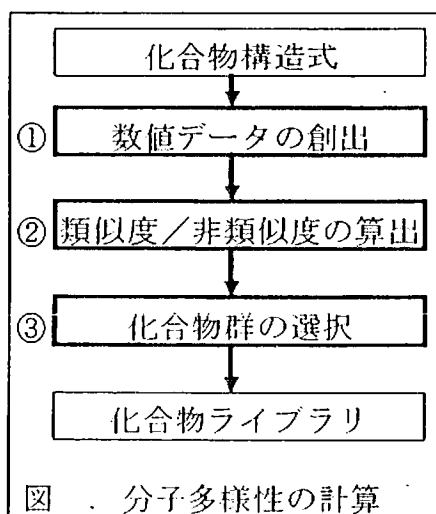


図 分子多様性の計算

分子多様性による化合物ライブラリ構築の目的は、多数の化合物群から構造的あるいは特性的に似た化合物群をまとめ、その中から一つ、あるいは少数の化合物を代表として取り出して、より少数の化合物群から構成される化合物ライブラリを構築する事である。

最初の化合物母集団から類似度／非類似度を求め、構造重複を許さずに特定の化合物群を選択して化合物ライブラリを構築する作業の標準的な流れを図に示す。以下、この流れ図に従って順に説明する。

①化合物構造式から数値データの創出

実際に分子多様性を実施する最初の手続きは、化合物構造式を一旦数値データ（パラメータ）に置き換えることである。この構造式を数値データに変換する手法の概要は、既に第 3 章のパターン認識法にて説明している。なお、Hansch-Fujita 法や 3-D QSAR においてもパラメータ創出に関する同様な記述があるが、これらの解析手法に利用されるパラメータはそれ自体が構造一活性相関に特化しているため、分子多様性評価での利用は特定の場合（例えばリード候補化合物ライブラリの構築等）に限定される。

パターン認識法で利用される様々なパラメータ群は、構造式が包含する様々

な情報を単純に数値データへと変換したものである。この点で、これらのパラメータは分子多様性にも利用出来る。実際に、トポロジカル、トポグラフィカル、物理化学的パラメータ等の総てが分子多様性の評価に利用可能である。これらの化学特性に相関したパラメータ群は、単純な構造情報だけに基づいた分子多様性の評価に化学的な特性に応じた評価を加味し、分子多様性解析に幅と深みを加える。

化合物構造のみを意識したパラメータとしては化合物の部分構造情報に基づくパラメータが良く利用される。この部分構造パラメータにも多数存在するが、分子多様性で良く利用されるパラメータは指定された部分構造が化合物中に存在する時を1、存在しない時を0とするバイナリ情報である。これは最も単純な部分構造パラメータであり、化合物検索の検索キーとして多くの実績がある。

また、1/0 情報であることは次項にて述べる類似度計算にも都合が良いという利点もある。

・フィンガープリントを用いた分子多様性の評価

Martin^{*1)}らにより実施された分子多様性評価で用いられたフィンガープリント (Fingerprint) は一種の部分構造パラメータである。名前の通りに、化合物の識別用 ID としての機能を目指して開発されたものであり、化合物構造式に特別の変換操作を適用することで創出される特殊なコードである。これは、小さな単位 (最大7結合) で構成された多数の部分構造に関する 2048 個の 1/0 情報からなるビット列 (bit string) である。

このフィンガープリント本来の開発目的は化合物の高速検索であり、分子多様性の評価を目指したものではない。化合物検索が目的であるが故に、構造式からフィンガープリントへの変換過程でコンピュータ技術にからむ特殊な操作 (ハッシング、データ圧縮等) を行っている。この結果、創出されたバイナリコードは比較的少ないビット数で化合物と1対1にユニークな対応関係を取るようになる (これがフィンガープリントの最終目的である) が、そのバイナリコード自体は構造式を正確に再現しているわけではない。つまり、指紋を見ただけでは元の人顔かたちが分からないのと同様、このビット列から元の構造式を再現することは出来なくなっている。この事実を考えるならば、化合物の類似性評価の元データとしてフィンガープリントだけを利用する時には注意が必要である。

* 1 : Eric J. Martin, Jeffrey M. Blaney, Michael A. Siani, David C. Spellmeyer, Alex K. Wong, and Walter H. Moos, *J. Med. Chem.*, 38, 1431-1436(1995).

②類似度/非類似度の算出

これらのパラメータ群を用いて化合物同志の類似度/非類似 (相違) 度を計

算する。この化合物間の類似度／非類似（相違）度は、パターン認識的にはパターン空間における化合物同志の距離関係に置き換えることが出来る。距離が0ということは二つのパターンが数値的に全く同一であることを意味する。このパターン間の距離を求める計算式は多数存在し、データの種類／内容等により適用される計算式が異なる。計算式が異なれば、例え同じデータを用いたとしても結果は多少異なってくる。従って計算手法の吟味が必要となるが、実務上は計算手法の差異よりも、むしろ用いるデータの吟味を行う方が効果は大きい。これは類似度に限らず一般解析でも当てはまる事である。少しでも良い相関係数や分類率を達成するために計算手法を変える事が行われているが、同じパターンを用いており、しかも同じ解析目的であるならば、計算手法自体の議論でない限り単に目先が変わるだけである。まず利用するパラメータを変えるのが先決であり、この方が正しい解析への近道である。

コンビナトリアルケミストリで最も良く利用される計算式が谷本の類似インデックス(Tanimoto's Similarity Index)である。

$$D_{\text{tanimoto}} = \frac{\sum_{i=1}^N [X_{ai}X_{bi}]}{\sum_{i=1}^N [X_{ai} + X_{bi} - X_{ai}X_{bi}]}$$

ここで、Dはパターンaとbの距離で、 X_{ai} および X_{bi} はパターンaおよびbのi番目のデータであり、値が1のもののみ計算する。 $X_{ai}X_{bi}$ は両方のi番目のパターンが共に1の時のみ1として計算する。従って、全次元のデータが全く同一の場合は1、総ての次元データが異なっていれば0の値を取る。この値を1から引けば、その値は非類似度を示すことになる。この谷本の類似インデックスはハミング距離(Hamming Distance)を改良したものであり、1/0のバイナリデータを用いるとき一般的に利用される。

また、1/0以外の数値データを用いる時は、以下の式で示されるミンコフスキー距離(Minkowski Distance)等が頻繁に利用される。

$$D_{\text{minkowski}} = \left[\sum_{i=1}^N (X_{ai} - X_{bi})^k \right]^{1/k}$$

この式において、 $K=2$ の時は一般的に利用されるユークリッド距離となる。この他、シティブロック距離、キャンベラ距離等があり、データの特性や状況に応じて使い分けられている。個々の詳細については専門書にゆずる。

・谷本の類似インデックスの計算例

谷本の類似インデックスがコンビナトリアルケミストリデータ利用される

大きな理由は、1/0 のデータにより表現されるパターン間の距離（類似度）を計算するのに最適という評価があるからである。以下に1/0 の17次元データを用いた、実際の計算例を示す。

表 . パターン a と b のバイナリデータの論理和と排他的論理和

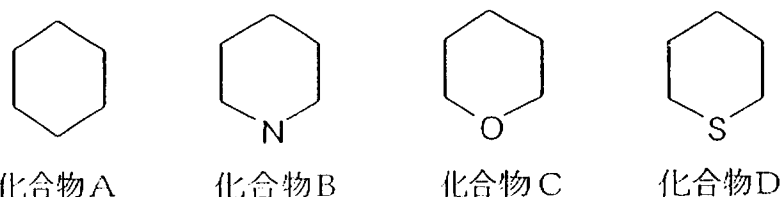
パターン a	1 0 0 1 1 0 1 1 1 1 0 1 0 0 1 0 0	
パターン b	1 1 1 0 0 1 0 0 0 1 0 1 0 0 1 1 1	
論理和 (OR)	1 1 1 1 1 1 1 1 1 1 0 1 0 0 1 1 1	• 14
論理積 (AND)	1 0 0 0 0 0 0 0 0 1 0 1 0 0 1 0 0	• 4

谷本の式において、分母は1/0 の二値データを用いた論理和 (OR) を、また分子は論理積 (AND) を意味している。表 には、パターン a とパターン b の各次元単位に、論理和と排他的論理和を取った結果が示されている。論理和の総和は14、論理積の総和は4 であるので、パターン a とパターン b との距離は0.286 となる。同様に、次元がどのように増大しようとも簡単にパターン a と b との距離を高速/的確に計算できるのが特徴である。

・バイナリコードを用いた谷本の類似インデックス計算の留意点

谷本の類似インデックスは単にビット情報の比較を行うだけである。従って、計算に用いるビット列の特性により類似度に差異が出ない事がある事を考慮しなければならない。

図 に示された化合物は化学的に異なる化合物群であるが、類似度は全く同じ値となる。部分構造の存在情報を用いることでビット列に違いが出る（化合物の違いは認識する）が、化合物間の類似度に差異は無い。通常はより複雑な構造を持つ化合物群を扱うのでこのような問題は顕在化しないが、データと計算手法が持つ本質的な特性を理解しておく事は必要である。



	C	N	O	S					
化合物A	1	0	0	0	1	0	0	0	1	SimAB = SimAC = SimAD = SimBC = SimBD = SimCD = 3/5 = 0.6
化合物B	0	1	0	0	1	0	0	0	1	
化合物C	0	0	1	0	1	0	0	0	1	
化合物D	0	0	0	1	1	0	0	0	1	

図 . 部分構造の有無によるビット表示と谷本の類似度計算結果

③化合物群の選択

類似度を求めたならば、この類似データに基づいてパターン母集団から類似

化合物群を取り除くことが必要である。例えば谷本のインデックスを求めただけではパターン同志の類似性は評価出来るが、どのパターンが同じグループ（多変量解析的にはクラスターと呼ぶ）に帰属するかを判定する事は出来ない。すなわち、パターン群を類似性の高いパターン群（クラスター）にまとめる手続きが必要である。この手続きとして、多変量解析のクラスタリング手法が利用される。例えば、階層的クラスタリング手法の出力であるデンドログラムを利用するならば、第 4 章でも述べたように横線で切断されたクラスターの数に最小パターン数が決定される。同様に、非階層的クラスタリングでは望むクラスター数になった時点で計算を止めれば良い。続いて、そのクラスターを代表する一つの化合物を選択する事が必要になるが、この場合はクラスターを構成する化合物群の重心に最も近い化合物を代表化合物とする。複数の化合物群を取り出す時は個々のクラスター内に均等に分散した化合物群を取り出す。個々のクラスタリング手法については第 4 章のパターン認識法に簡単な説明がある。

・ Jarvis-Patrick 法

ここでは分子多様性の評価に特異的に利用される Jarvis-Patrick 法について簡単に述べる。このアプローチは通常のクラスタリング手法とはかなりアプローチが異なっている。具体的には簡易型のクラスタリング手法であり、手続き的には分類手法である最近隣法（K-NN 法：第 4 章パターン認識法参照）に類似したアプローチである。最大の利点は計算時間の短い事があげられる。これは他のクラスタリング手法が全パターンが一つのクラスターになるまで、あるいは一つのクラスターが個々のクラスター（即ちパターン）になるまで何回も融合／分割が繰り返されるのに対し、本手法は限られた回数での繰り返し計算でクラスタリングを実施出来る為である。従って、通常のアプローチではパターン数が増大すると急速に計算時間が長くなるが、本手法ではパターン数に計算時間が大きく左右される事はない。

手続きは最近隣法と同様、最初に自分の周辺に近在する K 個のパターン群を決定する。取り出された K 個のパターンから一つ取り出し、そのパターンの自分を除く近隣パターンを K-1 個チェックする。取り出されたパターンの近隣パターンと自分の近隣パターンを比較し、同じパターンが J 個以上含まれている時は自分と取り出されたパターンは同じクラスターに帰属させる。J 個に満たない時は異なったクラスターに帰属させる。この操作を全パターンについて繰り返して実施する。なお、 $1 < J < K$ である。

現在、クラスタリング手法は多種多様の形で展開され、その数は多数におよぶ。著者も“超球概念”^{*1)}を導入した“超球クラスタリング”を提唱し、さらにファジィ理論を導入した“ファジィ超球クラスタリング”を展開し、構造

一活性相関研究に用い、従来のクラスタリング手法に比較して良い結果を出している*2)。クラスタリングに関する詳細な解説は専門書を参照されたい。

*1：湯田 浩太郎、計算機統計科学、

*2：湯田 浩太郎、情報化学討論会、

なお、先に述べた Martin らはフィンガープリントを用いて得られた谷本の類似インデックスのマトリクスデータを入力データとして、MDS(Multidimensional scaling)手法を用いて新たな5次元ベクトルデータへと変換している。この5次元ベクトルのうち、バリエーションの大きな順に取り出された最初の三ベクトルを用いて三次元プロットを行い、パターン群がクラスタ一化されていることを視覚的に捕らえている。

3.5 “リード候補化合物ライブラリ”構築

前節で述べた分子多様性による化合物ライブラリ構築の目的は、化合物の重複や偏りといった問題を純粹に化学構造だけの観点から追求し、サンプリングに起因する問題を避ける事である。従って、このライブラリ構築過程に構造一活性相関やドラッグデザインの考え/ノウハウは入っていない。このため、化合物空間を代表する化合物の絶対数は減少しているが、元の母集団におけるリード候補化合物の含有率はそのままである。

先にも述べたようにコンビナトリアルケミストリ/HTS と言えども、より少ない数の化合物でリード化合物が発見されればこれに越したことはない。特に合成を考えればこの問題は深刻である。

3.5.1 リード候補化合物ライブラリの特徴と二種類のアプローチ

理想的な化合物ライブラリとは、化合物の偏りや重複が無いと同時に、ライブラリ中にリード化合物を多数含んでいる事である。このように、リード候補化合物リッチという特性を持つ化合物ライブラリを“バイアスライブラリ”と呼ぶが、本著では“リード候補化合物ライブラリ”と称する。著者は以前より“リード候補化合物データセット”という概念を提唱していた事と、バイアスの意味する所は非常に大きく、内容的に様々なバイアスが考えられる事が主な理由である。ここでの目的はリード化合物の発見である事を考えれば、こちらの表現の方が適していると考えられる。

リード候補化合物の含有率が高められたリード候補化合物ライブラリの構築手段として、構造一活性相関で培われた技術が利用される。このリード候補化合物ライブラリの構築手続きとして、主に二種類のアプローチが考えられる。

①化合物創出を基本としたアプローチ

一つは構造一活性相関上の技術とノウハウを適用して化合物の部分構造を選択し、この選択された部分構造を用いて新たに化合物を組み立てるアプロー

チである。この創出された化合物群の分子多様性を評価した構造重複の無い化合物群を最終的なリード候補化合物ライブラリとする(図)。これは全く新規の化合物群を構築する時に取られる一般的なアプローチである。現在、化合物の組立に必要な部分構造の選択手法に対して様々な手法が提唱されている。これらの概要については次節にて述べる。

図中、背景が濃い色の四角部分が構造-活性相関に関係した項目である。実線の四角が現在実施されている部分で、点線で囲まれた四角は内容的に必要であるが現時点ではあまり行われていない手続きである。このように手続きが片手落ちであるのには幾つかの理由がある。この最大の理由は本著を読まれた読者であればすぐ気が付くと思うが、‘構造変化の大きな化合物群を対象とした薬理活性評価(予測)は、従来の殆どの構造-活性相関手法では実施出来ない’という事実が原因となっている。このような条件に対し、唯一適用可能な手法はパターン認識法のみである。

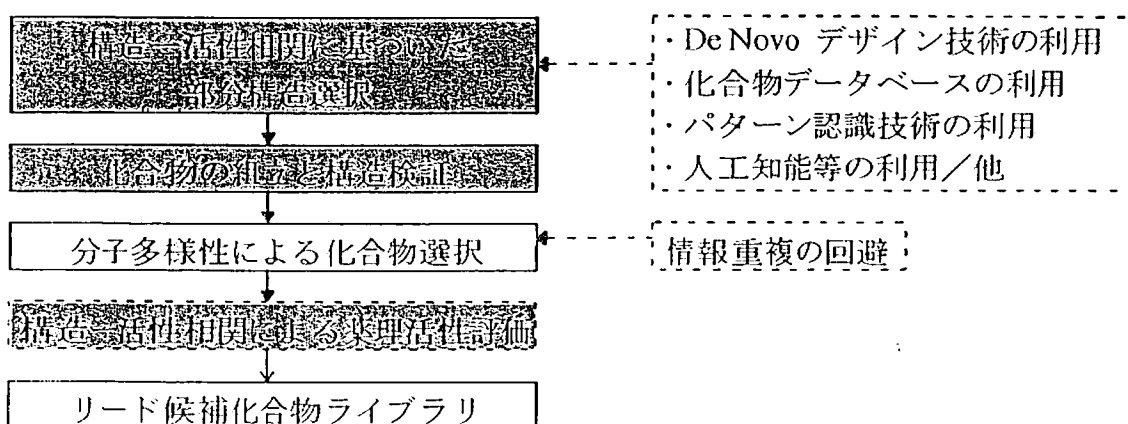


図 1. 部分構造の組立によるリード候補化合物ライブラリの構築

② 既存化合物群の評価/選択によるアプローチ

既存の化合物群(手続き的には既存/既知および未合成にこだわらない)について構造-活性相関上での評価を行い、その情報を基本としてリード候補化合物ライブラリを構築するアプローチである。これは、既存の化合物ライブラリを再評価してリード候補化合物リッチな化合物ライブラリとする場合に必要となる。例えば、試薬ライブラリや自社化合物ライブラリ等の見直し、実際にスクリーニングや合成を行う時の優先順位付け等に利用される。この手続きには既存化合物の活性評価機能が必要であり、この評価に適用出来るアプローチは構造の大幅に異なる化合物群を扱えるパターン認識法に限定される。図に本アプローチの流れ図を示す。

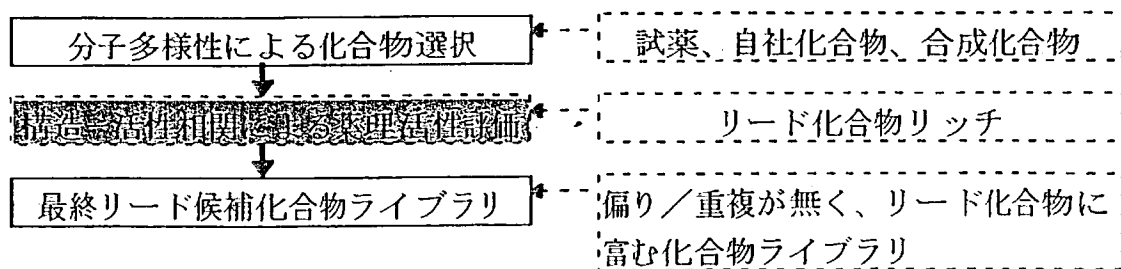


図 . 化合物選択によるリード候補化合物ライブラリの構築

なお、①/②いずれのアプローチにおいても、分子多様性による化合物選択の過程はこの一連の作業の中のどの位置に有っても結果を大きく左右する程の差異は生じない。

3.5.2 リード候補化合物ライブラリ構築に対する具体的アプローチ

①新規化合物群の創出を基本としたリード候補化合物ライブラリの構築

化合物の創出は化合物を組み立てるための部品（パーツ）を何らかの手法で集め、これらの組み合わせを行うことで実施される。この組み合わせ用部品の調達技術として構造-活性相関の技術やセンスが大いに利用されている。

現在行われている部品調達のための主要なアプローチは、De Novo デザイン技術を基本としたアプローチと化合物データベースを利用したアプローチの二種類である。これらの系列の他に著者が行った事例として、パターン認識法によるアプローチ（“リード候補化合物再構築”：第 3 章 De Novo デザイン参照）もある。さらに、人工知能技術を用いたバイオアナログ化合物群を創出するアプローチもパターン認識法と組み合わせ著者により実施されている。この人工知能による化合物創出の詳細については事例の項で詳しく説明する。

・ De Novo デザインを基本とするアプローチ

De Novo デザインで創出された化合物群は基本的に総てリード候補化合物ライブラリとなる。この De Novo デザインの詳細は第 3 章を参照されたい。コンビナトリアルケミストリーに De Novo デザイン関連技術を利用するアプローチは手続き的に大きく二種類ある。一つは、化合物組立用の部品と基本構造をレセプターサイトとのフィッティング過程でどのように選択してくるかという選択を重視したアプローチで、この選択手法の差異によりシステムとしての特徴が出る。

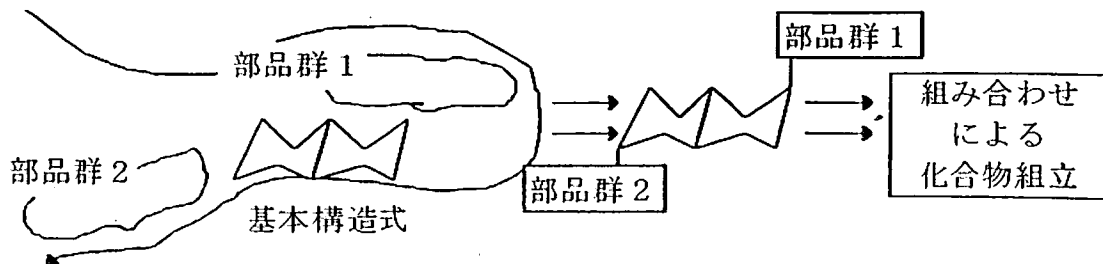


図 . 部品群のフィッティングによる選択と組み合わせによる化合物創出

もう一つは、ファーマコフォア等の三次元的な相対的位置関係をフィッティング等により切り出し、これを基本に三次元検索キーを構築して三次元化合物検索と連動させ、取り出された化合物の基本構造や部分構造を用いて新たな化合物群を組み立てるアプローチである。

相対的特性の取り出し 三次元検索キーの作成

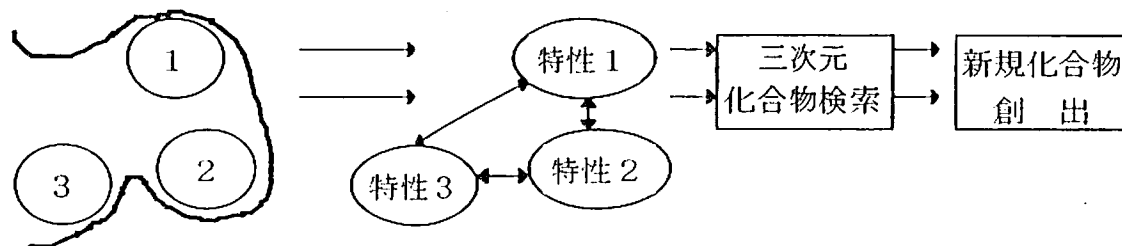


図 . グループ間の相対的位置関係の取り出しと三次元化合物検索との連携

・本アプローチの一般的特徴

基本骨格を固定する事も出来るが、一般的に本手法により創出される化合物群の構造的な差異は大きい。この故に、新規リード候補化合物としての価値は高い。構造的に大きく異なる新規化合物群の合成という厄介な問題を除けば、スクリーニングする価値は高いと考えられる。しかし、De Novo デザインの章でも述べたように自由度が非常に高いという本質的な特性から、創出された化合物の薬理活性に対する信憑性は他のアプローチと比較して低いと考えざるを得ない。少なくとも、充分条件としての薬理活性評価機能と組み合わせる事が必要である。

・化合物データベースを利用するアプローチ

この種のライブラリ構築で最も単純なアプローチは化合物データベースを利用するアプローチである。本アプローチの具体的な手続きを図 に示す。以下、この図の順に従って説明する。

・最初に目標活性を有する化合物群を文献や化合物データベース中から複数取り出す。

・この取り出された化合物群から、共通となる基本骨格を抜き出す。この基本骨格が決まったならば、置換基の位置や種類についてその変化パターンを取り出す。

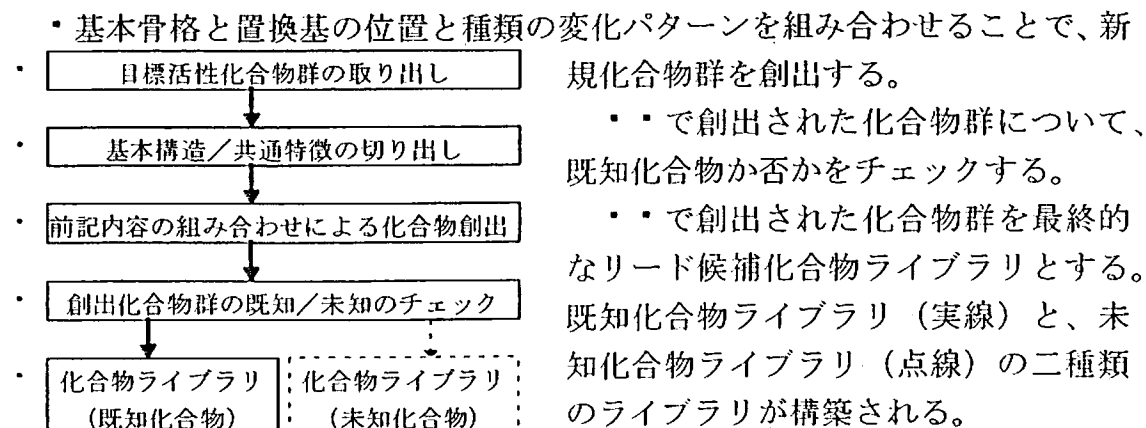


図 化合物データベースを利用した化合物ライブラリの構築

最も簡単な例としてサルファ剤関連化合物群から導かれる置換基変化を例に取り説明する。・の目標化合物群の取り出しの後、それらの化合物群から切り出された基本骨格と置換位置の特定 (この場合は R_1 と R_2)、及び置換基変化が図 に示される。この段階で、化合物創出のための条件が揃ったことになる。単純な組み合わせを行うことで、 5×5 の 25 種類の化合物が創出される。この時、置換基の種類をバイオアイソステリズム (Bioisosterism: 生物学的等価性、藤田は Bioanalogy^{*1)} と称している) の考えに従って増やせば簡単に化合物数は増大する。

*1: 第 3 章、人工知能による構造-活性相関 (EMIL)

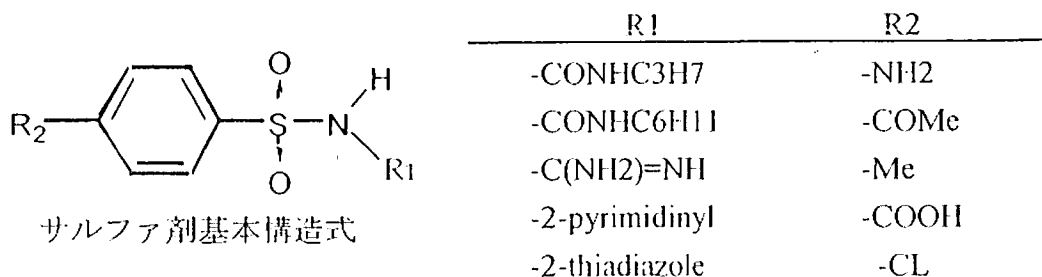


図 サルファ剤の基本構造から導かれた薬物群の置換基変化

・本アプローチの一般的特徴

創出される化合物群は基本骨格が固定され、単に置換基の変化のみのため、本アプローチによる化合物ライブラリの構造的な相違は小さいのが特徴である。従って、合成も他の手法で創出された化合物群と比較してかなり容易である。このアプローチでは感度が低いアッセイ系を用いると殆ど総ての化合物群が活性となる事が多い。これは、基本的に同じレセプターサイトに反応する活性化合物群の基本骨格および官能基を雛形としており、且つ創出された化合物

群と雛形化合物との構造的差異が小さいためと考えられる。

以上の特性を考えた場合、本アプローチは新規リード化合物の発見というよりも特許対策や見落とし防止策としての価値の方が高い。

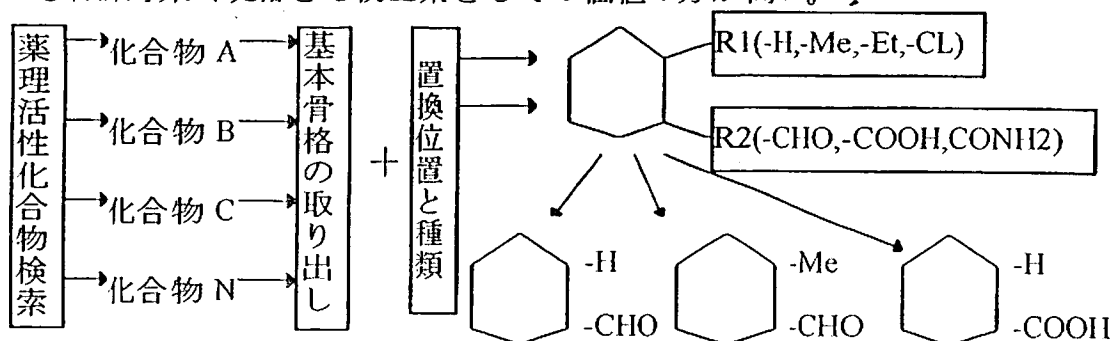


図 ． 化合物データベースを利用した新規化合物の創出

・パターン認識によるアプローチ

パターン認識を適用することで新規のリード候補化合物ライブラリを構築する事が可能である。これは、パラメータとして部分構造パラメータを用いる事で達成される。この手法の詳細は第 4 章の De Novo デザインにおける適用事例の中で、二次元 De Novo デザイン（“リード候補化合物再構築”）として紹介してある。

このアプローチの最大の特徴は、創出された化合物群についてパターン認識法により改めて薬理活性の予測を行っていることである。つまり、リード候補化合物としての必要条件と充分条件を備えた化合物が最終リード候補化合物ライブラリとして選択される。

・本アプローチの一般的特徴

創出される化合物群は基本骨格が同じで、かつ置換位置も同じであるが、置換基は選択された総ての部分構造同志や基本骨格と結合し、位置特異性の無い事が特徴である。このために、
・ 程の構造変化はないが、
・ 以上の構造変化はある。
本手法では他の手法と異なり薬理活性評価を行っており、リード候補化合物ライブラリとしての信頼性は高い。しかし、位置特異性のない部分構造情報を用いて新たに化合物群を組み立てる為、構造要求が厳しいメカニズムを持つ薬理活性への適用は困難が予想される。

・人工知能等の利用／その他のアプローチ

著者は人工知能システムの EMIL とパターン認識解析システムである ADAPT を連動させ、新規化合物を創出すると同時に五種類の薬理活性を評価する試みを実施している。この詳細については本章の事例の項で改めて述べる。

・本アプローチの一般的特徴

EMIL により創出される化合物群は基本骨格の変化を含めたもので、構造的変化は大きい。しかも構造変換を複数回適用するとこの変化は更に大きくなり、

創出化合物数も急激に増大する。この為、ADAPT を用いて創出された化合物群の薬理活性予測を行い、更に信頼性の高いリード候補化合物ライブラリの構築を実現している。他の手法と比較してEMILは簡単な操作で新規リード候補化合物群を多数創出する。しかも、初期化合物を指定出来るので合成し易い化合物等を出発化合物として展開出来る利点がある。しかし、化合物変換に用いるルールの内容の違いにより創出化合物の構造式が大きく変化し、変換ルールの数が少ない時はあまり変化しないという特徴を持つ。

これらのアプローチの基本は、化合物組立に必要な部品を選択する過程で構造-活性相関の技術やセンスを応用することである。このように、部品を切り出す過程は構造-活性相関の基本原理に基づくが、組み合わせて創出された化合物群に対する活性からの評価や構造-活性相関的な考察は著者が行ったアプローチ以外殆ど行われていないのが現状である。以下に前記四種類のアプローチの特徴を簡単にまとめる。

構造-活性相関手法	部品切り出しの基本	化合物創出上の問題
De Novo デザイン	→ フィッティング	<ul style="list-style-type: none"> ・ 組み合わせの爆発 ・ 部品の差異による構造式の差が大 ・ 創出化合物の活性保証無し
データベース利用	→ 過去の活性化化合物	
パターン認識法	→ 部分構造パラメータ	
人工知能の適用	→ 過去の構造変換事例	

②既存化合物群の選択によるリード候補化合物ライブラリの構築

新規化合物の合成も魅力あるが、既存の化合物群を利用する事は合成の観点や自社内化合物の再評価や、他社との差別化の観点から重要である。このアプローチは化合物の創出が目的ではなく、化合物の選択が目的である。

・相対的な位置関係を基本とするアプローチ

特定の官能基（ファーマコフォア等）群が三次元的に相対的な位置関係を保つ事が薬理活性の発現に重要な役割を果たすことはドラグレセプター理論の基本から明白である。この事実を利用して活性化化合物群を選択する試みは三次元化合物検索等での基本であり、同時にこの技術はコンビナトリアルケミストリのリード候補化合物選択にも適用可能である。但し、通常の検索と異なる点は単なる取り出しのみならず、取り出された化合物群の多様性評価機能が必要な事である。なお、ファーマコフォアを利用して分子多様性を求めるアプローチは一般的にファーマコフォア多様性解析 (pharmacophore diversity analysis) と呼ばれる。

◎三点ファーマコフォア法の適用

Davies らは Y.C.Martin らが提唱した三点ファーマコフォア距離 (three-point pharmacophore distance)^{*1, 2)} に基づく解析を実施した^{*3)}。彼が用いたファーマコフォアは、水素結合供与基、水素結合受容基、チャージ (+) 中心、芳香族環の中心、疎水性中心、酸性中心、塩基性中心の7種類である。これらのファーマコフォアから三個取り出すのに 84 通りの組み合わせがある。また、各ファーマコフォア間の距離は 0~15Å の距離を 31 段階に分けた。この結果、 $31^3=29791$ 通りの三点トライアングルが理論的に構築可能である。実際は物理的な制約 (三角形の最長辺が他の二辺の和よりも長い時) によりこれよりも数は少ない。先のファーマコフォアと距離関係を組み合わせると、総計 2489556 個の三点トライアングルが形成される。各トライアングルを 2496000 個のビット列に置き換え、これをファーマコフォアキー (Pharmacophore key) とした。各化合物についてこれらのトライアングルの有無を確認し、存在する時 1、存在しない時 0 としてファーマコフォアキーを作成した。この操作より化合物群は先のフィンガープリント同様のビット列に置き換えられた事になる。分子多様性の評価はこのビット列を利用して行う。

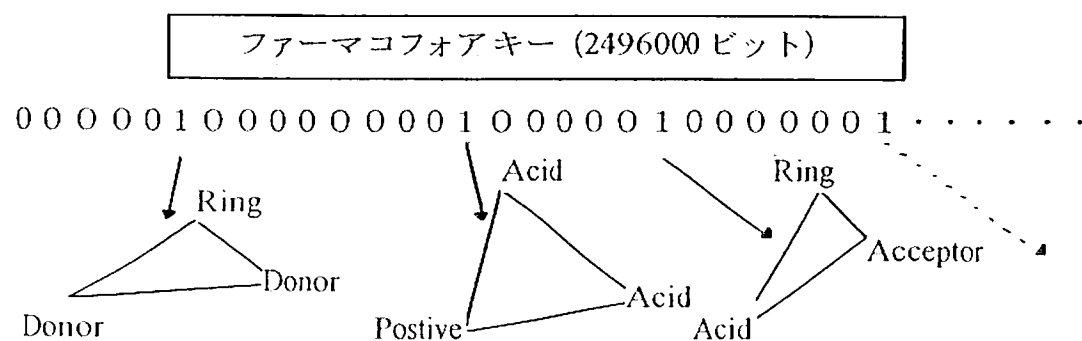


図 . 個々の化合物に設定されたファーマコフォアキー

このようなアプローチは構造-活性相関的には極めて自然であり、受け入れ易い考えである。

* 1 : Y.C.Martin, et al., J.Comput.Aided Mol. Des., 2, 15 (1988).

* 2 : Y.C.Martin, "In Random and Rational Drug Discovery via Rational Design and Combinatorial Chemistry", Strategic Research Institute, New York (1995).

* 3 : Keith Davies, 'Using Pharmacophore Diversity to Select Molecules to Test from Commercial Catalogues', in "Molecular Diversity and Combinatorial Chemistry (Libraries and Drug Discovery)", Eds., Irwin M. Chaiken and Kim D. Janda, pp.309-316, American Chemical Society, Washington, DC (1996).

・パターン認識によるアプローチ

これは著者により行われたアプローチである。本アプローチの目的はプレ

スクリーニング用の化合物ライブラリを化合物データベースの中から選択する事である。このアプローチの詳細は第 4 章のパターン認識法にて“リード候補化合物選択”として、事例の項で述べている。本手法の特徴は構造特異性の少ないパラメータ群で構成された判別関数を用いている事である。これにより構造変化に富んだ化合物群（例えば一般化合物データベース等）の活性予測が可能となる。他の手法が物性等の一般的な基準を用いた分子多様性に基づく分類にしか過ぎない事を考えるならば、薬理活性予測を行い事実上のプレスクリーニングを行う本アプローチの重要性は高い。分子多様性に基づいて選択された化合物群の評価等に本アプローチを適用することで、効率の良いスクリーニングを実現する事が可能となる。

・その他のアプローチ

多数の既存化合物群を高速に、且つ簡単にフィルタリングするには幾つかの基準を設けてこれらの条件を満たす化合物群を取り出す事である。この判定に用いられる指標としては構造-活性相関で重要なパラメータである LogP、分子屈折率、電子的パラメータ、分子量等の物性値を用いる。従来、このような物性値は解析に利用する目的で求められていたが、今後は化合物の簡単なフィルタリングを目的とした利用が増えてくると考える。

単純なフィルタリングのみならず、実際のライブラリ構築の為には分子多様性の評価が必要となる。Mason らは 6 種類の物性および関連特性をそれぞれ 2~3 のクラスに分類し、これらを組み合わせることで化合物群を分類している^{*1)}。このアプローチは、各次元が 2~3 のクラスデータからなる 6 次元空間上に化合物群をプロットした事に相当する。この場合の分子多様性は物性を基準とした分子多様性であり、化合物の構造式に基づいた外面的な基準に基づく分子多様性ではない。

* 1 : J.S.Mason, I.M.McLay, R.A.Lewis, P.M.Dean, G.Jolles, C.G.Newton, Eds., Academic Press, London, p.225 (1994).

Martin らは 6 種類の特性（原子半径、酸性、塩基性、水素結合供与性、水素結合受容性、芳香族性）情報を用いて化合物を特徴付けしている^{*1)}。この特性情報は置換基の置換位置からの距離（結合数で考える）に従って各原子について計算され、この場合は 15 原子まで計算している。従って、各化合物は 6 種類×15 個の数値データから構成される。このデータを用いて一種のミニマックス法（Min-max technique）的考えを適用し、以下の式を用いてパターン間の類似度を求めている。

$$Sim_{AB} = \frac{\sum_{j=1}^6 \left(\sum_{i=1}^{15} D_{ij}^A \wedge \sum_{i=1}^{15} D_{ij}^B \right)}{\sum_{j=1}^6 \left(\sum_{i=1}^{15} D_{ij}^A \vee \sum_{i=1}^{15} D_{ij}^B \right)}$$

式中 Sim_{AB} はパターン A と B の類似度で 0~1 の値を取る。値が 1 の時は、パターン A と B は全く同じ特性を持つ。 D_{ij}^A および D_{ij}^B はそれぞれパターン A および B の i 行 j 列の特性値を示す。記号 \wedge はその記号の左右の値のうち小さい方の値を取ることを、 \vee は大きい方の値を取ることを示す。

* 1 : Eric J. Martin, Jeffrey M. Blaney, Michael A. Siani, David C. Spellmeyer, Alex K. Wong, and Walter H. Moos, *J. Med. Chem.*, 38, 1431-1436(1995).

3.5.3 化合物ライブラリ構築手法の比較

現在、分子多様性に基づく化合物ライブラリやリード候補化合物ライブラリのいずれに対しても、様々なアプローチが試みられつつある段階にある。どのアプローチが決定/理想的であるかの結論は最も興味ある所ではあるが、実際にコンビナトリアルケミストリ/HTS で展開された新規リード化合物に関する情報が少ないという現状では極めて困難である。

評価というものは簡単なようで非常に難しい。例えば、フィンガープリントもファーマコフォアキーも共に形式は 0/1 で構成されるビット列である。但し、情報の中身はフィンガープリントが化合物構造情報を基本としたビット列であるのに対し、ファーマコフォアキーは受容体マップを強く意識した機能中心の情報である。形式は同じでも、中味は根本的に異なっており、実際の作業はこの中身によって結果が大きく異なってくる。フィンガープリントは情報化学分野で展開されてきた歴史を持ち、ファーマコフォアキーは構造-活性相関解析での実績を基本としている。どちらがコンビナトリアルケミストリに適しているかの検証は困難である。

データ一つ取ってもこのように大きな差異がある。コンビナトリアルケミストリの実施にはこれ以外の様々な要因が複雑に関与しており、簡単にどの手法が最適であるということ結論付ける事は現在の段階では危険と考える。現時点で手法間の比較を行った報告として Brown らの実験がある^{*1)}。彼らは計算に用いる七種類のパラメータ群、四種類のクラスタリング手法、および四つの化合物セットを用いて比較を行っている。彼らの報告では化合物検索用に展開された重み付き情報が有効で、クラスタリング手法は Ward 融合手法を用いた結果が良いとの結論が導かれている。繰り返すが、新しい手法が出ると必ず比較問題が展開される。構造-活性相関においても新たな手法が出ると旧アプローチと比較する論文が出る。新規の手法を提案する時に行う比較は必要である

が、第三者が行う時は開発者以上に本質を見極めたという確証がなければ危険である。さもなければ真に公平な比較は出来ないと著者は考えている。

・* 1 : R.D.Brown, Y.C.Martin, J. Chem. Inf. Comput. Sci., 36, 573 (1996).

4. リード候補化合物ライブラリデザインの問題点と活性予測機能の必要性

4.1 リード候補化合物ライブラリが抱える三種の問題点

本著を読まればすぐわかるが、化合物ライブラリを創出/選択するのに必要な技術は全く新しいものではなく、従来から展開されてきた構造-活性相関技術の範囲内で充分に対応出来る事項である。例えば、部分構造取り出しとその組立に至る一連の流れは De Novo デザインそのものであり、既存化合物群からの選択も三次元化合物検索が果たす役割は大きい。また、パターン認識により、リード候補化合物群を化合物データベースから取り出す“リード候補化合物検索”、活性に重要とみなされる部分構造をパターン認識により特定し、新たに化合物群を構築する“リード候補化合物再構築”も既に著者により行われている。以下に、現在行われているリード候補化合物ライブラリが抱える一般的な問題点を三点程あげる。

①リード候補化合物であるという化合物ライブラリの信頼性保証の問題

コンビナトリアルケミストリ/HTS 全体の流れを考えた時、部分構造の組み合わせで自動的に構築された新規化合物群が本当にリード候補化合物となりえるか否かの判定は、ライブラリの信頼性を評価する意味で極めて重要である。なぜならば、創出される化合物群の絶対数が多く、個々の化合物の構造変化が大きく、しかもその殆どが新規化合物であるという三重苦を考えるならば、合成上の観点からもこの問題は深刻である。当然ながら絵に描いた餅だけでは新薬に結びつかない。

残念ながら、現在行われているアプローチの殆どはリード候補化合物の創出機能のみであり、その活性予測は行っていない。つまり、化合物を組み立てる時に利用される部分構造の決定に構造-活性相関上の技術やノウハウを十二分に活用しているが、一旦創出された化合物に関し、その活性予測は行っていないか、その予測機能は機能的に貧弱というのが現状である。これを例えるならば、車を作るのに必要な部品は吟味して揃えるが、これらの部品を組み合わせで出来上がった製品は何のテストもしないで売り出す事と同じである。活性を示すための必要条件のみを満たし、充分条件を満たす事なくリード候補化合物ライブラリデザインを行っている。少なくとも製品調査に該当する活性予測機能は化合物ライブラリには必須の機能である。

また、ライブラリ構築時に限らず、HTS を実施した後に活性化合物群が多数残るような場合には二次スクリーニングにかけるための順位付けが必要とな

る。このような順位付けも一種の予測であり、パターン認識法を適用する事で可能となる。

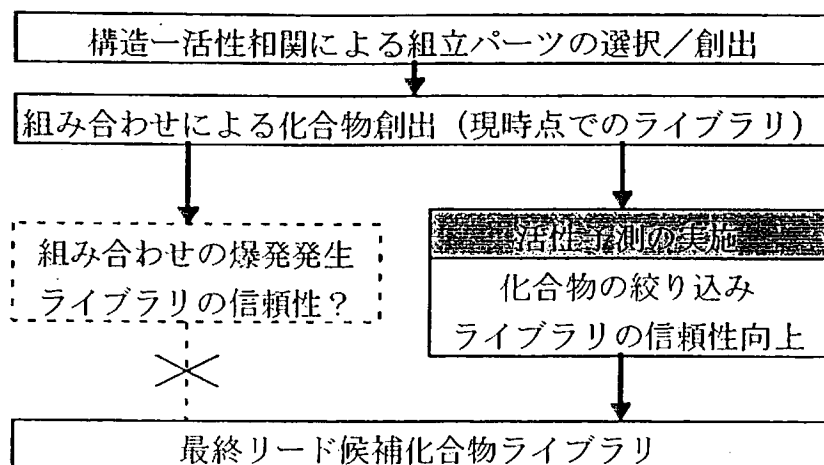


図 . リード候補化合物ライブラリにおける活性予測の必要性

②組み合わせの爆発に起因する問題

組み合わせの爆発の項でも述べたように、たとえ少ない数のパーツを用いたとしても創出される化合物数はコンビナトリアルケミストリ/HTS の処理限界を簡単に越えてしまうという問題がある。まさに、数に驕る者は数に泣くのである。このため、組み合わせに用いる部分構造を組立以前の選択段階で可能な限り絞り込むことが必要となる。しかし、リード候補化合物が化合物空間中に均一に分散していると仮定するならば、化合物創出する前の過程で絞り込む場合と、総て創出した後に絞り込む場合とでは大きな差異が出る。一般的には、少ないパーツを用いて組み立てられた製品中に目的製品が含まれる絶対数は相対的に小さくなる。一方、多くのパーツを用いて組み立てた場合は製品数も大きくなるが目的製品の数も増える。従って、リード候補化合物群を多く残すならば、化合物創出後に化合物群を選択することが望ましい。この事実は、①の問題と深く関与している。

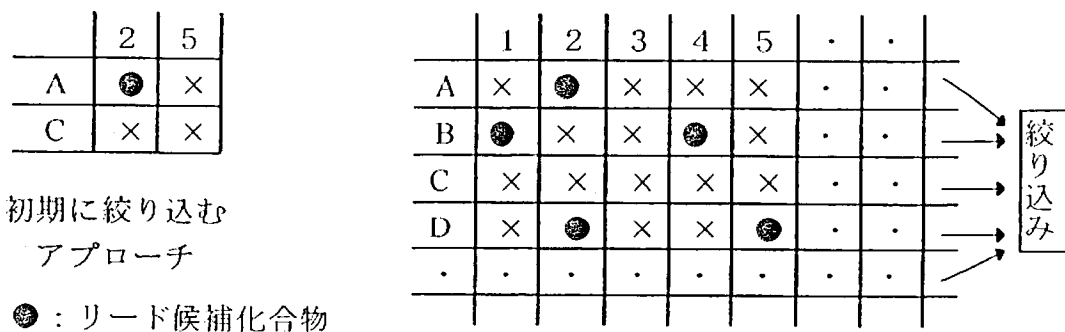


図 . 初期に絞り込むアプローチと創出後に絞り込むアプローチ

③創出される化合物の変化性の問題

純粹に技術的な問題であるが、化合物創出には必ず自由度の問題がつきまとう。既に De Novo デザインの章にても述べたが、如何に工夫しようとも創出される化合物をユニークに決定することは不可能である。つまり、例え同じ酵素（レセプターサイト）を用いたとしても、使うシステムの差異や、適用するオプションの差異、研究者の主観の差異等の様々な要因により最終的に創出される化合物群の構造式は大きく異なってくるという事実がある。創出構造式が変わっていてもそれらがリード候補化合物であれば問題はない。しかし、このような検証が困難であるという事実と、油断すると創出化合物数が膨大なものになり、合成も困難という事を考えると、実用上の観点で不安が残る。

ここでは、現在一般的に行われている化合物ライブラリの構築手法に関する三種類の問題点をあげた。このように、いかに構造-活性相関の技術やノウハウを適用しようとも、活性予測機能による化合物の評価と絞り込みが出来なければ、現在のリード候補化合物ライブラリは絵に描いた餅となる。

ここであげた問題点は、創出された化合物群を効率よく活性予測する事ができればある程度片づく問題である。作業の上流過程で如何にノイズが紛れ込もうとも、下流過程でこのノイズを取り除くか減少させる事が出来るならば作業全体として何ら不都合はない。コンピュータにおける永遠の課題である“組み合わせの爆発”や“自由度”の問題に正面切って取り組むよりも、何らかの評価機能を付加する事の方が実用的である。以下ではこの化合物ライブラリの薬理活性予測問題について論じる。

4.2 化合物ライブラリの活性予測における特徴と適用可能な手法の種類

①化合物ライブラリの特徴（分子多様性が高い）

コンビナトリアルケミストリ/HTS で利用される化合物ライブラリの最大の特徴は化合物構造の多様性である。分子多様性の技術を駆使して既存の化合物データベースから選択された化合物群であり、また新たに創出された化合物群にしても分子多様性は十分に考慮されている。

②分子多様性が高い化合物群を対象とする活性予測法

構造-活性相関の各章でも述べたように、活性予測が可能な手法は限定される。ドラグレセプター理論によるアプローチと De Novo デザインは単にレセプターサイトとのフィッティング度合いによる評価であり、直接的な薬理活性予測は出来ない。さらに、薬理活性予測が可能な構造-活性相関手法のうち、Hansch-Fujita 法はその原理上基本構造や置換位置が固定されており、構造的な要求が最も厳しい。3-D QSAR はHansch-Fujita 法と比較すると扱える化合物の自由度は高いが、同一レセプターサイトでフィッティング可能な構造範囲が原理上の適用限界となる。

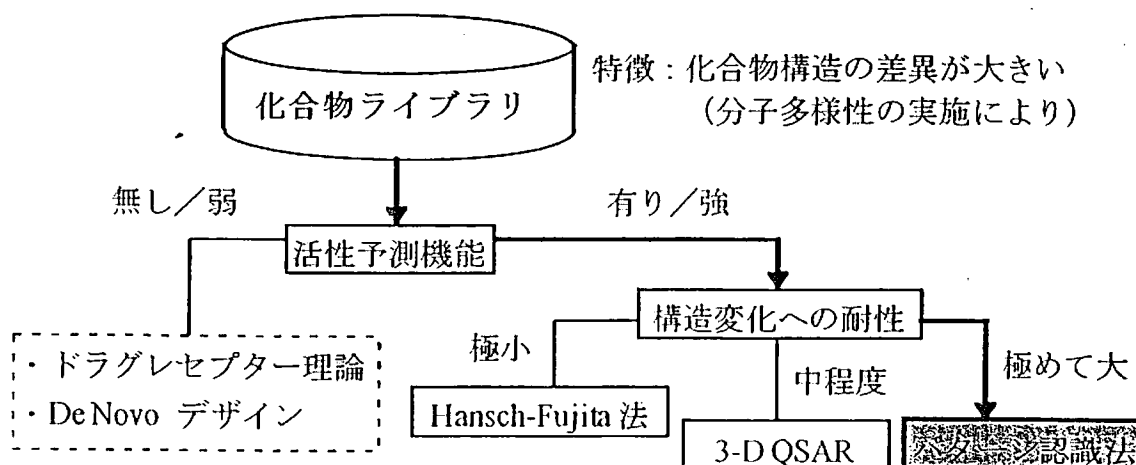


図 ． 化合物ライブラリに対する活性予測の適用図

パターン認識法による解析の基本は対応関係の利用である。この対応関係を構造制限の小さい/パラメータと薬理活性との間に設定する事で、構造的に多様な化合物群の活性予測が可能となる。De Novo デザインでも一部で活性予測を行なうシステムもあるが、この活性予測に関する手続き部分はパターン認識法を利用している。従って、現時点で構造変化の大きな化合物ライブラリの活性予測はパターン認識法でしか実現されていない。図 ． に化合物ライブラリと活性予測における種々の構造-活性相関手法との関係を示す。

化合物構造式	制限内容	構造-活性相関手法
完全に異なる化合物群	構造制限無し	パターン認識法
異なる化合物群	フィッティング可能な範囲	3D QSAR
同族体化合物群	基本骨格/置換位置固定	Hansch-Fujita 法

③活性予測を行ったリード候補化合物ライブラリの構築

著者は以前、プレスクリーニング手法として“リード候補化合物検索 (Lead retrieval)” を提唱している。これは、化合物データベースから無作為に取り出された全く異なる構造を持つ化合物群をパターン認識法を用いて活性予測を行うもので、そのまま化合物ライブラリの活性予測に適用可能である。このリード候補化合物検索は第 ． 章のパターン認識法の事例の項にて詳しく述べてある。

また、新規化合物のライブラリ構築についても同様にパターン認識法を利用した“リード候補化合物再構築 (Lead re-construction)” を提唱している。こちらは、化合物の組立に利用する部分構造の選択をパターン認識を用いて行い、基本構造式と部分構造を用いて新規化合物を組み立て、この新たに創出された化合物群を判別関数を用いて活性予測している。

前記いずれのアプローチにおいても化合物の活性予測を行っており、現在提

唱されている化合物ライブラリ構築手法の殆どが薬理活性予測機能を持っていない事を考えるならば、その有用性は高い。“リード候補化合物検索 (Lead retrieval)” は第 4 章のパターン認識法にて、また“リード候補化合物再構築 (Lead re-construction)” は、第 5 章の De Novo デザインにおいて二次元 De Novo デザインとして説明されている。

先にも述べたように、パターン認識法による活性予測は単に化合物ライブラリの構築時のみならず、HTS の一次スクリーニングにおいて多くの化合物が活性と判断された時の二次スクリーニングの優先順位付けにも利用可能である。

4.3 化合物創出と化合物選択によるリード候補化合物ライブラリ構築手順

実際に HTS が稼働し始めると化合物ライブラリの重要性が急速に高まる。この化合物ライブラリの構築には先にも述べたように化合物を創出するアプローチと、既存化合物から選択するアプローチとの二種類ある。既に前節までにこれらのアプローチについて述べてきたが、簡単にその内容をまとめてみる。

図 1 に化合物を創出して化合物ライブラリを構築するアプローチを示す。

先ず化合物群を創出するアプローチとして現在は、①化合物を構成する部品群を決定し、それらの部品群を用いて新規化合物群を創出するアプローチ、および②入力された化合物を出発化合物とし、この化合物を連続的に変形する事で

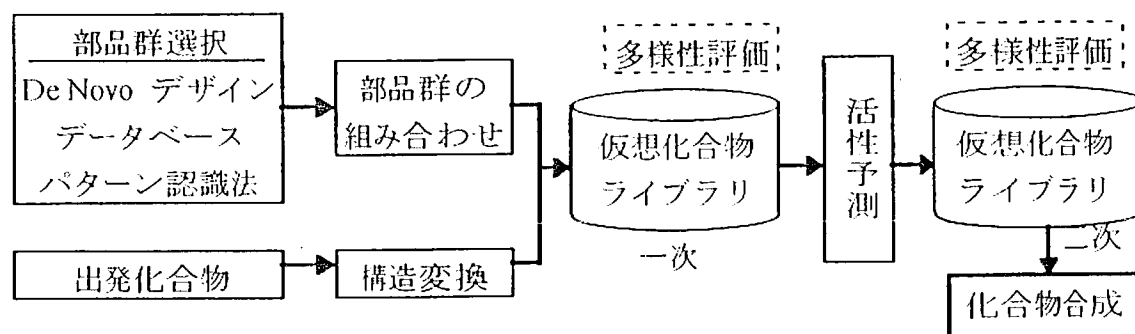


図 1 化合物構造創出による化合物ライブラリの構築

新規化合物群を創出するアプローチの二種類のアプローチに分けられる。①のアプローチは、De Novo デザインや化合物データベースを利用したアプローチと、著者が行ったパターン認識法によるアプローチがある。②のアプローチとしては本章で事例として取り上げるが、藤田らが開発した EMIL システムを用いて化合物群を創出するアプローチがある。このような化合物を修正して新規化合物群を創出するアプローチは本アプローチのみである。

本アプローチの最大の特徴は、出来上がった化合物ライブラリは殆どが仮想化合物群であり、最終的に合成することが必要となることである。この合成が実用上最大のネックとなる。創出された化合物群は仮想一次化合物ライブラリとなり、活性予測を行わない時はこの時点で分子多様性の評価が必要となる。

この仮想一次化合物ライブラリレベルでは先に述べた合成の観点からまだまだ絞り込みが不十分である。従って、この仮想一次化合物ライブラリについて更に活性予測を行い、より信頼性が高く、化合物数の小さな化合物ライブラリとしたものが仮想二次化合物ライブラリである。現時点でこの活性予測が可能なのは、原理的に構造変化の大きな化合物を扱えるパターン認識法による識別関数を用いた予測しかない。

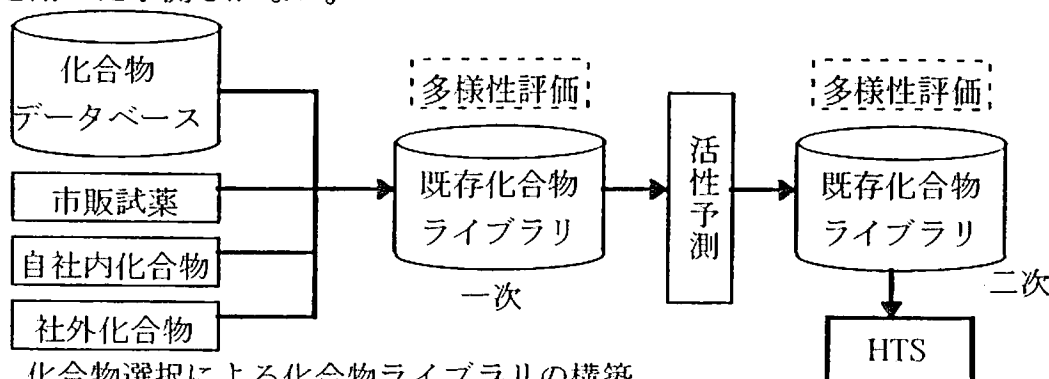


図 化合物選択による化合物ライブラリの構築

図 に既存化合物群から選択して化合物ライブラリを構築する時の作業の流れをまとめた。このアプローチでは化合物として既存のものを扱う、従って化合物データベース、市販試薬、自社内化合物および社外の研究機関等で合成された化合物群を集めることが必要である。これらの化合物群について分子多様性を評価したものが既存一次化合物ライブラリとなる。また、前項同様さらに信頼性の高い化合物ライブラリ構築には活性予測が必要となり、この操作を経た化合物群が既存二次化合物ライブラリとなる。但し、既存化合物群を扱う場合は化合物群が HTS の処理数を大きく越えていない限り総て HTS にかけるのが原則と著者は考える。

5. 活性予測を行ったリード候補化合物ライブラリ構築事例

前節でも述べたように、著者はパターン認識法を応用することで薬理活性予測を伴ったリード候補化合物ライブラリの構築について既に二種類のアプローチを展開してきた。さらには、人工知能による化合物創出支援システムである EMIL システムにより創出された化合物群をパターン認識により活性予測を行い、“バイオアナログ化合物ライブラリ”の構築を試みている。薬理活性予測を行った化合物ライブラリの構築はコンビナトリアルケミストリ/HTS のプレスクリーニングとして実用的観点から重要な意味を持つ。

5.1 活性予測を行ったバイオアナログ化合物ライブラリの構築*¹⁾

本アプローチはバイオアナログ化合物群を EMIL システムにより創出し、その創出された化合物群の薬理活性の予測を ADAPT システムの判別関数を用いて行いバイオアナログ化合物ライブラリを構築する。EMIL および ADAPT システムの連携により、コンビナトリアルケミストリ/HTS のプレスクリーニ

ングを実現させる。

* 1 : 湯田 浩太郎、“バイオアナログ化合物ライブラリを用いた、バーチャルコンビナトリアルケミストリ/HTS の実施”、第 24 回構造-活性相関シンポジウム講演要旨集、P.309、大阪、1996 年。

5.1.1 実験概要

実験の目的：五種類の薬理活性について活性予測されたリード候補化合物（新

規）ライブラリを構築する。

解析の流れ：EMIL システムデータ創出されたバイオアナログ化合物群を五

種類の薬理活性について活性予測し、活性と判断された化合物をライブラリに加える。

実験手続き：・ EMIL システムに出発化合物を入力する。

・ 医薬品変換ルールを用いて、出発化合物の構造変換（一次変換のみ）によりバイオアナログ化合物群を創出する。

・ 予め ADAPT 上で五種類の薬理活性について判別関数を用意しておく。

・ EMIL にて創出されたバイオアナログ化合物群を ADAPT

シ

システムに移行し、五種類の判別関数に該当するパラメータ群を創出する。

・ 創出されたパラメータ群を用いて、五種類の薬理活性を予測する。活性有りとした化合物をライブラリに加える。

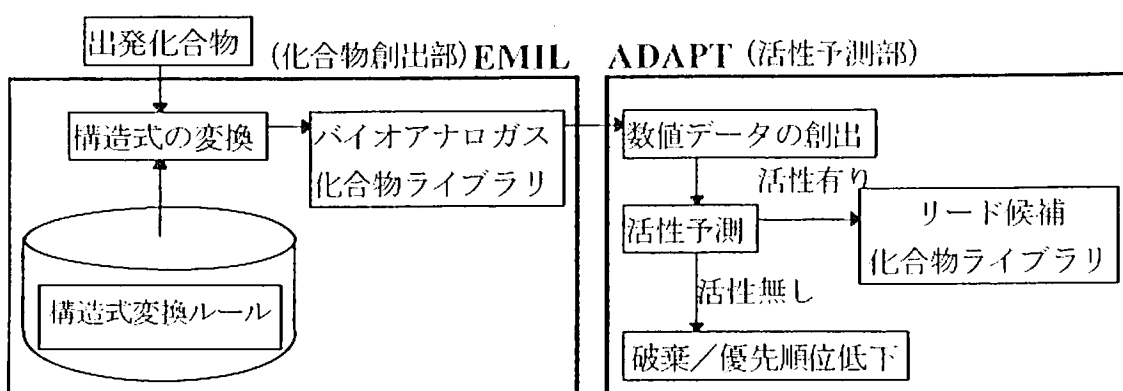


図 . Emil/ADAPT 連携によるリード候補化合物ライブラリの調整

5.3 作業の流れ

本アプローチは最初にバイオアナログ化合物群を EMIL により創出し、その創出された化合物群に対して ADAPT システムの判別関数を用いて活性予測

を行うものである。ADAPT により活性有りとされた化合物が最終リード候補化合物ライブラリとなり、活性無しとされた化合物は破棄するか、優先順位を低くする。なお、活性予測に用いられる判別関数は予め解析を行った上で、ADAPT に保存しておく事が必要である。

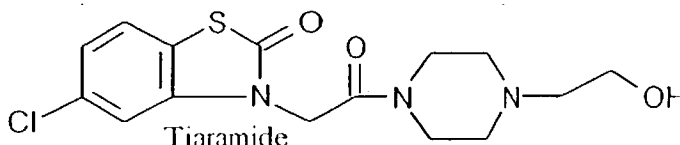
① EMIL システムによるバイオアナログ化合物の創出

EMIL システムは人工知能によるドラッグデザイン支援システムである。システムには予め過去に行われたリード化合物構造変換事例が構造変換ルールとして保存されている。次に、入力された化合物を出発化合物とし、保存されている構造変換ルールを順次適用することでバイオアナログ（リード候補）化合物群を創出するシステムである。本システムの詳細は第 4 章を参照されたい。

EMIL には初期化合物として図 1 の化合物（tiaramide）を入力した。この入力された化合物に医薬品用の化合物変換ルールを適用した。第一次変換として 78 個の化合物群が出力された。これらの化合物群は薬理活性変換ルールの基本となった薬理活性とバイオアナログな特性を持つ。更に第二次、第三次、・・・と構造変換を続けることで化合物は多数創出されるが、今回は第一次変換の 78 化合物とその一次変換化合物から取り出した一化合物を出発構造式として創出された二次変換化合物の 76 化合物の、総計 154 化合物に tiaramide を加えた 155 化合物の活性予測を行った。

① 出発化合物構造式

Antiasthmatic,
Anti-inflammatory



② 適用した構造式変換ルール

③ 化合物変換結果

第一次変換出力構造式数 78 化合物

第二次変換出力構造式数 76 化合物

（一次変換化合物中の一化合物を出発化合物として利用）

図 1 . Emil で用いた出発化合物と化合物変換情報

図 2 に EMIL システムにより創出された一次変換化合物の一部を表示する。

図 3 . EMIL システムにより創出された一次変換化合物群

② 活性予測のための判別関数の取得

活性の予測は判別関数を用いて行うが、この判別関数は第 4 章でのパターン認識法による解析から得られる。今回は ADAPT システムを用いて判別関数を創出した。予め用意された判別関数は五種類で、その薬理活性と判別関数取得に用いられた化合物の種類は以下のようになる。

- ・ CCK-A (Cholecystokinin-A) 受容体拮抗剤 急性膵炎治療薬
 - ・ Malonamide 誘導体
- ・ Ca²⁺遊離抑制作用 降圧薬、抗狭心症薬
 - ・ 1,5-Benzothiazepine 誘導体
- ・ 抗腫瘍作用
 - ・ 4-Bis [(aminoalkyl) amino] -9,10-anthracenediones 誘導体
- ・ 抗菌作用
 - ・ Quinolone および 1,8-Naphthylidine 誘導体
- ・ 抗アレルギー作用
 - ・ Imidazo 誘導体

③活性予測のためのパラメータ創出

EMIL により創出されたバイオアナログ化合物群全 155 化合物を ADAPT システムに転送し、前記五種類の薬理活性予測に必要となるパラメータ群を創出する。これにより薬理活性の予測が可能となる。

④判別関数によるバイオアナログ化合物群の活性予測

ADAPT により予め用意されていた五種類の判別関数を用いて EMIL により創出された 155 個のバイオアナログ化合物の活性予測を行う。この活性予測結果を表 に示す。

表 . 判別関数による五種類の活性予測

化合物 ID	活性 1	活性 2	活性 3	活性 4	活性 5	化合物 ID	活性 1	活性 2	活性 3	活性 4	活性 5
1	×	○	○	○	○	79	×	○	○	×	○
2	×	○	○	○	○	80	×	×	○	○	○
3	○	×	×	×	×	81	×	×	×	○	○
4	×	×	○	○	○	82	×	○	○	○	○
5	×	×	○	○	○	83	○	×	×	×	×
6	×	×	○	○	○	84	×	×	×	×	×
7	×	×	×	○	×	85	×	×	×	×	○
8	○	×	○	○	×	86	×	×	○	○	○
9	○	×	○	○	○	87	○	○	○	○	○
10	×	×	○	○	○	88	×	○	○	○	○
11	×	×	○	×	○	89	○	×	×	×	×
12	○	×	○	○	○	90	×	×	×	○	○
13	×	×	○	○	○	91	×	×	○	○	○
14	×	○	○	○	○	92	×	○	○	○	○
15	×	○	○	○	×	93	×	×	○	○	×
16	×	○	○	○	○	94	×	×	○	○	○
17	○	○	○	○	×	95	○	○	○	○	○

18	○	○	○	○	×	96	×	○	○	×	○
19	○	○	○	○	○	97	×	○	×	○	○
20	○	○	○	○	×	98	×	×	○	×	○
21	×	×	○	○	○	99	×	×	○	○	○
22	×	×	○	○	○	100	×	×	○	×	×
23	○	○	○	○	×	101	×	×	○	○	×
24	○	○	×	○	○	102	×	×	○	○	○
25	○	○	○	×	○	103	○	×	○	○	×
26	×	×	○	○	○	104	○	○	×	○	×
27	○	○	○	×	○	105	○	○	×	×	○
28	○	○	○	○	○	106	×	×	○	○	○
29	×	×	○	○	○	107	×	×	×	×	○
30	○	×	○	×	○	108	×	×	○	○	○
31	×	○	○	○	○	109	○	×	×	×	○
32	○	×	×	×	○	110	×	×	○	○	×
33	×	×	○	○	×	111	×	×	×	×	×
34	○	○	○	○	○	112	○	○	○	○	○
35	×	×	○	○	×	113	○	○	○	○	○
36	×	×	○	○	○	114	×	○	○	○	○
37	○	×	×	×	○	115	×	×	○	×	×
38	○	×	○	×	×	116	×	×	○	○	○
39	×	○	○	○	○	117	×	×	○	○	×
40	×	○	○	○	○	118	○	○	○	×	○
41	○	○	○	×	×	119	×	×	○	○	○
42	×	×	○	○	○	120	○	○	×	×	○
43	○	○	○	×	○	121	×	×	×	×	×
44	×	×	×	○	○	122	×	×	○	○	○
45	×	○	×	×	○	123	×	×	○	○	○
46	×	×	○	○	○	124	○	×	×	×	×
47	×	○	○	×	×	125	×	×	×	○	○
48	○	×	○	○	×	126	○	×	×	○	○
49	×	○	○	○	○	127	×	○	×	×	×
50	○	×	×	○	○	128	×	×	×	○	○
51	×	○	○	×	×	129	×	×	○	○	○
52	○	×	×	○	○	130	×	○	○	○	○
53	×	×	○	○	○	131	×	×	×	○	○
54	×	×	×	○	○	132	○	○	○	○	○
55	○	○	○	○	○	133	×	○	×	×	○
56	×	○	○	×	○	134	×	×	×	×	×
57	○	○	○	○	○	135	×	○	×	×	○
58	×	×	×	×	×	136	○	○	○	○	×
59	×	○	○	○	○	137	○	×	×	×	×
60	×	×	○	×	○	138	○	×	×	×	○
61	○	×	○	×	○	139	○	×	×	×	×
62	○	○	○	×	×	140	○	×	×	×	×
63	○	○	×	×	×	141	○	×	×	×	×

64	○	○	×	×	×	142	×	×	○	○	×
65	×	○	○	○	○	143	×	○	○	○	○
66	×	×	○	×	×	144	○	○	×	×	○
67	×	○	○	×	○	145	×	×	○	×	○
68	○	×	×	×	○	146	○	×	×	×	○
69	×	×	○	○	○	147	×	×	○	○	○
70	×	×	○	○	○	148	○	○	×	○	○
71	×	×	○	○	×	149	×	×	○	○	○
72	○	○	×	○	○	150	×	×	×	×	○
73	×	×	○	×	×	151	×	×	×	×	○
74	×	×	○	×	×	152	×	×	○	×	×
75	×	×	×	×	○	153	×	×	×	○	○
76	×	○	○	○	○	154	○	×	×	○	○
77	×	×	○	○	○	155	×	×	○	×	○
78	○	×	×	○	○						

活性 1	活性	57 化合物	非活性	98 化合物	活性化合物割合
36.8%					
活性 2	活性	59 化合物	非活性	96 化合物	活性化合物割合
38.1%					
活性 3	活性	103 化合物	非活性	52 化合物	活性化合物割合
66.5%					
活性 4	活性	96 化合物	非活性	59 化合物	活性化合物割合
61.9%					
活性 5	活性	109 化合物	非活性	46 化合物	活性化合物割合
70.3%					

⑤考察

EMILにて創出された155のバイオアナログ化合物を用いて、五種類の活性についてその予測を行った。薬理活性の種類により活性と判断された化合物の割合が36.8%から70.3%まで大きく変動した。さらに個々の化合物について活性と判断された割合を見ると、総ての薬理活性に活性と判断された化合物(10化合物)から、総て非活性と判断された化合物(5化合物)まで多岐に渡ることがわかる。三種類活性と判断された化合物が52化合物と最も数が多い。図に総て活性と判断された10化合物と、総て非活性と判断された5化合物を合わせて表示する。最初の二行が活性化合物で、最後の一行が非活性化合物群である。

これらの化合物群を見て活性及び非活性に両極端に分類されたかを簡単に述べる事は困難である。特に、活性と評価された387、412および432番と総てに非活性とされた411、421そして434との関係は簡単には言葉にならない。

どれだけ正しいのかは実際の実験で証明するしかないが、少なくともコンピュータはこれらの構造式間でこれだけの差異を認識したことは事実である。

⑥まとめ

今回行ったアプローチはパターン認識法の適用によりコンビナトリアルケミストリ/HTS のプレスクリーニングとしての役割を担う事が可能であることを示した物と考えている。活性に有効と見られる部分構造等を調達し、その組み合わせで多数のリード候補化合物群を創出することはコンビナトリアルケミストリの分野のみならず De Novo デザインの分野でも既に行われている試みである。これらのアプローチは必要条件のみを満たす化合物群を創出する為に、より信頼性を高めるための充分条件を満たすアプローチは取られていない。充分条件を完全に満たすためには実際に実験を行うしかない。しかし、その実験が総ての化合物について行えない、あるいは優先順位を付けたい場合には今回行ったような活性予測を行う事が重要な作業となる。パターン認識法の中でも述べたが、構造式の変化が大きい化合物群の活性予測はパターン認識でしか出来ないのが現状である。

ここで行ったように企業等では複数の薬理活性について同時に活性予測をする事は殆ど考えられないであろう。しかし、このように複数の薬理活性の評価が可能ということは、目標とする薬理活性の他にその副作用や毒性等の評価も可能ということの意味している実際の開発に着手する前に、薬理活性の他に副作用や毒性等をも踏まえた評価が出来ることは効率的な新薬開発には欠かせないものとなるはずである。

化合物の順位付けは、単に化合物ライブラリの調整のみならず、HTS を行った後のリード候補化合物群の開発順位付けにも役に立つ。特に同族体等を扱う HTS では活性と判断される化合物の数も多くなる。これらのリード候補化合物群を実際の薬物へと導くためには、ドラッグデザインによりさらに効果が高く、副作用や毒性の無い薬物へと展開する必要がある。この手順は複雑で手間のかかる作業であり、展開される化合物数には限界がある。この点でも化合物の順位付けは極めて重要な作業である。

6. 組み合わせ化合物 (Combinatorial Compounds) の考えと提案

コンビナトリアルケミストリ/HTS に適用される低分子化合物は非常に限定される。これは、変化に富む一連の化合物群を簡単に合成するアプローチが少ないためである。現在実用化されている化合物群としては、1,4-Benzodiazepine

ライブラリ^{*1)}、Mercaptoacyl proline ライブラリ^{*2)}、 β -Lactam ライブラリ、

Sulfamoylphenyls ライブラリ、Steroidal ライブラリ、 β -D-Glucose ライブラリ、Glycopeptide ライブラリ等がある。

*1 : Jonathan A. Ellmann, et. al., J. Am. Chem. Soc., 114, 10997 (1992).

*2 : Mark A. Gallop, et. al., J. Am. Chem. Soc., 117, 7029 (1995).

このように現時点のコンビナトリアルケミストリ/HTSにおける合成は、最初に化合物(ライブラリ)有りきの考えで展開されている。このような展開をする限り、総ての化合物を揃えるためには膨大な仕事量と時間が必要である。そこで化合物中心から、目的中心の考えに変更し、コンビナトリアルケミストリ/HTSを効率よく実施し、かつリード化合物の発見効率を上げるための化合物(組み合わせ化合物: Combinatorial Compounds)はなにか、という観点から考えて見る。

6.1 合成手法からのアプローチと、化合物からのアプローチの差

コンビナトリアルケミストリ/HTS業務で実施される合成概念と、従来から行われてきた合成概念とは、基本的な要素が180度変化する。この変化の内容について、化合物の量、ターゲット化合物の数、合成手順1および2の四項目について以下にまとめる。

表 . 合成に対するアプローチの違い

	コンビナトリアルケミストリ/HTSの合成	従来からの合成
化合物の量	微量	大量
構造変化量	大(分子多様性)	小(誘導体中心)
目標化合物	極めて多種類	一種類
手順1	平行合成	段階的合成
手順2	複数/同時合成	一化合物合成
ターゲット構造式	必要無し	必須
構造決定	活性評価後	ステップ毎に必要
化合物の純度	強くこだわらず	高純度要求

表からもわかるように、コンビナトリアルケミストリ/HTSの合成と従来からの合成とは殆ど総ての点で基本が異なってくる。この他にも目的の違いによる手続きの違い等も大きな差異を生む。例えば、従来の合成ではターゲット化合物の構造式が必要であったが、コンビナトリアルケミストリ/HTSでは必ずしも必要ではない。また、合成過程での化合物構造式が不明であっても大きな問題ではない。さらには、途中のプロセスにおける化合物の純度も大きな問題とはならない。これらの際だった特徴は、活性が見いだされた時点で構造式を決定しても構わないからである。つまり、実験が正確に再現されて、構造式等が後からでも決定出来るならば途中過程での構造決定に必要な労力を一切

省く事が出来る。

このような、基本項目自体の変化に対応する手段として二通りのアプローチが考えられる。一つは合成手法のみを変化/改良するもので、ターゲットとなる化合物に対する概念は従来のままである。固相合成の展開やロボットの利用等が本アプローチに該当し、従来型の化合物群を現在の技術範囲でいかに効率よく合成するかという、合成の目から見たアプローチである。この場合、目的としての化合物に対する概念の変化は無い。一方で、コンビナトリアルケミストリ/HTS を化合物から見て議論する事も可能である。すなわち、コンビナトリアルケミストリ/HTS を効率良く実施する時にあるべき化合物の特徴はどのようなものか、という観点からライブラリ構築について考える。

コンビナトリアルケミストリ/HTS における合成の、微量、多種類、平行/同時合成のいずれの特徴を取っても、百年以上にも及ぶ歴史に支えられた従来型の合成手法にはなじまない。これらの特徴を合成の立場から見るのではなく、化合物の立場から眺めた時新しい化合物群の存在が見えてくる。つまり、前記条件を満たすコンビナトリアルケミストリ/HTS を実施し易い化合物であること。同時に、リード化合物になる可能性の高い化合物であることが大切である。これらの多彩な要件を満たす化合物の形の一つとして“組み合わせ化合物”、および“組み合わせ試薬”を提唱する。

6.2 “組み合わせ化合物”と“組み合わせ試薬”

コンビナトリアルケミストリ/HTS で要求される化合物の特徴について考えると、①前節で述べたような様々な特徴を持つ化合物であること、②分子多様性の高い化合物群を簡単に合成出来ること、③薬理メカニズムに基づいた基本構造（リード化合物となる可能性が高い）を持つ事がある。これらの三条件を満たす一つの答えが組み合わせ化合物/試薬である。

組み合わせ化合物の基本は、合成の単位となる三種類の構造ブロックが連結したものである。

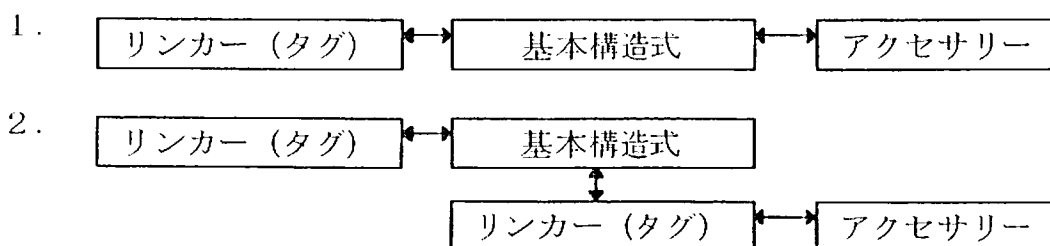


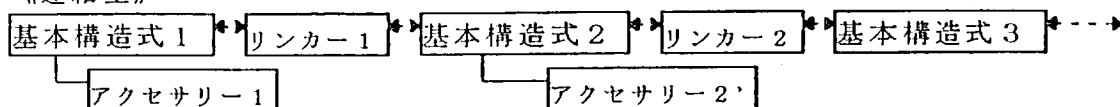
図 . 組み合わせ化合物の基本型

構造ブロックは、・基本構造式、・リンカー（タグ）、・アクセサリーの三種類存在する。基本構造式は化合物の骨格を構成するものであり、この基本構

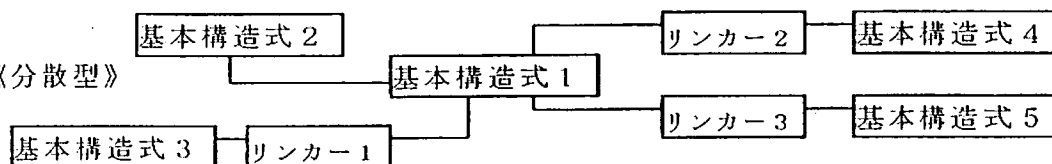
造式にリンカーやアクセサリが結合する。リンカーは基本構造式同志の結合や、基本構造式とアクセサリとの結合を介在する結合ユニットである。アクセサリは置換基レベルの構造を持ち、基本構造式と直結するか、リンカーを介して結合する。

これらの三構造単位が結合するパターンとして連結型と分散型がある。

《連結型》



《分散型》



連結型は基本構造式が直鎖状に結合しており、ペプチド的な構成を持つ低分子化合物である。分散型は、一つの基本骨格に数個の基本骨格が結合し、丁度火花が爆発して発散するような基本構造を持つ。

6.3 薬理活性と“組み合わせ化合物”

化合物は基本的に薬理活性が期待出来る（リード候補化合物）事が必要である。新たに提唱する組み合わせ化合物はドラグレセプター理論におけるリガンド化合物のフィッティング仮説を基本とする。つまり化合物が薬理活性を示すにはレセプターサイトへのフィッティングが必要であり、リード化合物は必然的にレセプターサイトにフィットするにふさわしい構造を取る。組み合わせ化合物もこの事実を満たす事が出来れば活性発現に必要な条件を満たす事になる。

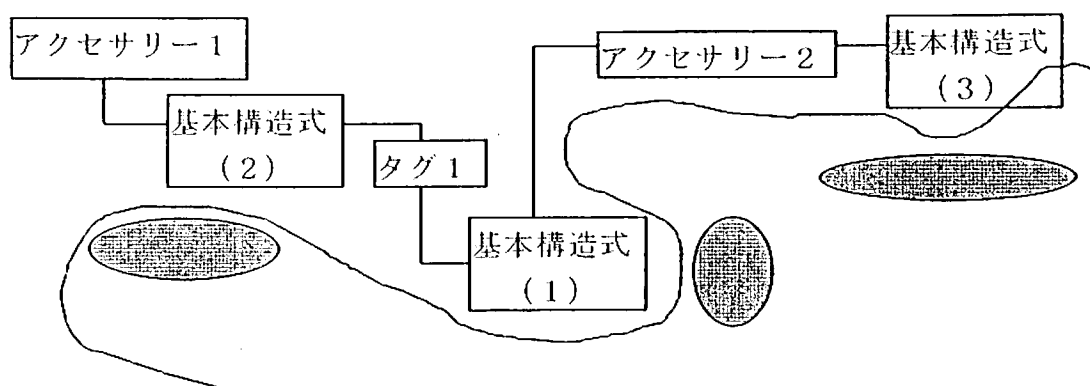


図 . レセプターサイトにフィッティングする組み合わせ化合物のイメージ

鎖状構造をなす組み合わせ化合物は酵素上に点在する様々な活性部位にフィットし易い形を持つ。最近になりこのような鎖状構造を持つ低分子化合物が

薬理活性を持つ報告が多くなっており、既存の薬物にもこのような特徴を持つ化合物が存在する。このような事実は、鎖状（側鎖ではない）の化合物は薬理活性を示す可能性が高いことを意味している。また、一般低分子化合物としてはこのような形状を取る化合物は少ない。

6.4 コンビナトリアルケミストリ/HTSにおける合成

6.4.1 “組み合わせ試薬”の導入

ここでは“組み合わせ試薬”をあらたに導入する。この組み合わせ試薬は、前項で述べた組み合わせ化合物を構成する、①リンカー（タグ）、②基本構造式、③アクセサリーの三種類の総称とする。

コンビナトリアルケミストリ/HTS に要求される化合物の合成は多数の化合物群を簡単に合成することである。どのような化合物が活性を示すか分からないのであるから、目的を決めた合成の必要はない。多数で、構造的に多様性のある化合物群を重複なく高速合成出来れば良い。さらに、プロセスが明確で再現可能であれば、化合物は混合物でも構わないし、副生成物であっても構わない。構造式は活性有りとなされた時点で決定されれば良い。

ターゲット指向合成から目的指向合成への転換！

このような従来からの合成の常識を越えた要求は、“組み合わせ試薬”を用いた合成で達成可能である。更に、このような組み合わせ試薬による合成はコンピュータによる合成設計薬理活性スケジューリングが容易になるという長所がある。従来からの合成では合成時にターゲットデザインが必要であるし、合成するためのデザインも必要である。構造が大きく異なる化合物群の合成は、試薬も反応条件もステップも総ての点で異なり、このような合成は、ロボットやコンピュータに取り、最も扱いにくいタイプの仕事である。

組み合わせ化合物をターゲットとするならば、化合物合成は少数の組み合わせ試薬を用いてロボットにセットするだけとなる。これで分子多様性を満たす多数の化合物群を高速に合成することが可能となる。組み合わせ試薬を合成するか、市販の組み合わせ試薬を用いるかして、自動合成ロボット等により圧倒的多数の化合物群を自動的、かつ効率良く合成する事になる。しかも、従来の化合物群とは構造的に大きくかけ離れ、化合物同志の分子多様性を高度に保ち、真のリード化合物を高密度で含んでいる。化学者はあいた時間を探索研究に使う事が可能となる。

組み合わせ試薬による合成	従来型の合成
試薬を合成する	化合物自体を合成する
試薬の組み合わせによる合成	個別単位での合成
構造式決定は活性評価後に行う	合成途中に構造決定が必要
ターゲットデザイン必要無し	ターゲットデザイン必要
合成デザイン必要無し	合成デザイン必要

6.4.2 化合物合成から組み合わせ試薬合成の勧め

化合物をターゲットとした合成では、1個の化合物を合成しても1個である。つまり、50個の化合物を合成しても50である。従来の合成手法は個々の化合物合成が対象であり、化合物を合成した時点で総ての作業は完結する。このような作業形態は一個の生産を完璧にする伝統工業的生産手法であり、数の大きさを競う場合の生産手法には適していない。

一方、試薬を合成した時は状況が一変する。単純のために基本構造式とアクセサリーだけに限定するならば、リンカー部分が三カ所ある基本構造式を20種類、アクセサリーを30種類合成すれば、理論上 $20 \times 30^3 = 540000$ 化合物合成したのと同じ効果がある。例え、1000化合物につき1化合物しか合成出来ないにしても540化合物は簡単に合成出来る。複雑な化合物の合成には合成計画も難しく、失敗の可能性も高く、時間もかかる。試薬であればその構造は比較的簡単であり、その分だけ前記項目の負荷は大きく減少する。研究者の負担を減らし、その分実のある研究に時間を割く事が可能となる。

このように合成目標を限定する必要が無く、数が勝負の合成には化合物単体の合成ではなく、試薬合成が最も合理的と著者は考える。この試薬合成を技術として定着させるにはさらに様々な検討を加えることが必要である。

6.5 化合物合成スケジュールのためのシステム

コンビナトリアルケミストリ/HTSで行われる合成は従来からの合成と異なる特徴を満たす事が要求される。比較的構造の異なる化合物群を多数、出来れば同じような実験条件で、高速に合成しなければならない。実際にコンビナトリアルケミストリ/HTSを行うと合成の前のプランニングに多大な時間を取られてしまうのが一般的な事象となる。合成ロボット自体も進化し、容器やウエル単位で合成条件をコントロールすることが可能となっている。しかし、いかにロボットが高性能になろうとも、合成条件を考えるのは研究者である。個々の反応単位で合成条件をロボットに入力するのはいかにも効率が悪すぎる。この意味からもコンピュータ支援による合成プランニングシステムが必要である。

単一の化合物をいかに効率よく合成するか¹の訓練を受けてきた研究者にと

り、多数の化合物の合成条件を相互にバランスを取りながら効率の良い合成計画を立てることは極めて困難である。この点でパラレル合成のプランニングを支援するシステムの存在が重要となるが、現時点でこのような合成プランニング支援システムは存在しない。今後の早急なる展開が望まれる。

6.6 化合物ライブラリと合成ロボット

紙上の化合物ライブラリを実際の合成に結びつけて現実のものとし、HTSに導く為のステップとしては合成ロボットとの連携が必用となる。この作業は従来のような一つの反応だけを管理する合成とは異なり、多数の試薬と反応条件の平行/同時管理が必用となる。ここでもコンピュータの助けが必用となる。

現在展開されている合成ロボットでは、多段階の合成をこなし、ターゲット化合物を純粋な形で取り出すに至る全行程を自動化する事は困難である。この意味で現実的には合成自体もロボットの限界から大きな制約を受ける事になる。

つまり、組み合わせで理論的に創出される化合物群と実際のコンビナトリアル合成には以前として大きなギャップが存在する。実際の合成にあたっては、先にも述べたようなコンビナトリアルケミストリでの合成と通常合成の基本的な違い、合成ロボットの限界等を考慮した上での合成手続きを考える事が必用である。

現在の合成管理支援システムはその内容から大きく二種類に分類する事が可能である。一つは、①コンビナトリアルケミストリ的思考を基本に、網羅的な化合物ライブラリ創出優先のアプローチ。残るアプローチは、②従来の合成的指向から出発してコンビナトリアルケミストリ的思考を取り込んだ反応パターン/試薬優先のアプローチである。

組み合わせ優先のアプローチは情報化学的発想より展開されており、正にコンビナトリアルケミストリの本質的アプローチといえる。しかし、コンピュータ的に可能性の総てを尽くす事がそのまま化合物合成に直結するとは限らない。組み合わせ出来ても合成出来ない/困難な化合物は多い。組み合わせで創出された化合物群について、新たに研究者が試薬や合成手法についてプランニングする事が必用である。

反応/試薬優先のアプローチは、先ず反応パターンとその反応に用いる試薬を先に考える。続いてその試薬の組み合わせにより可能な化合物を創出する。このような手順を踏めば、創出される化合物数は組み合わせ優先アプローチよりも減少するが、創出された化合物群は基本的に合成可能であり、且つ合成設計する必用はなくそのまま合成ロボットに仕込む事が可能である。

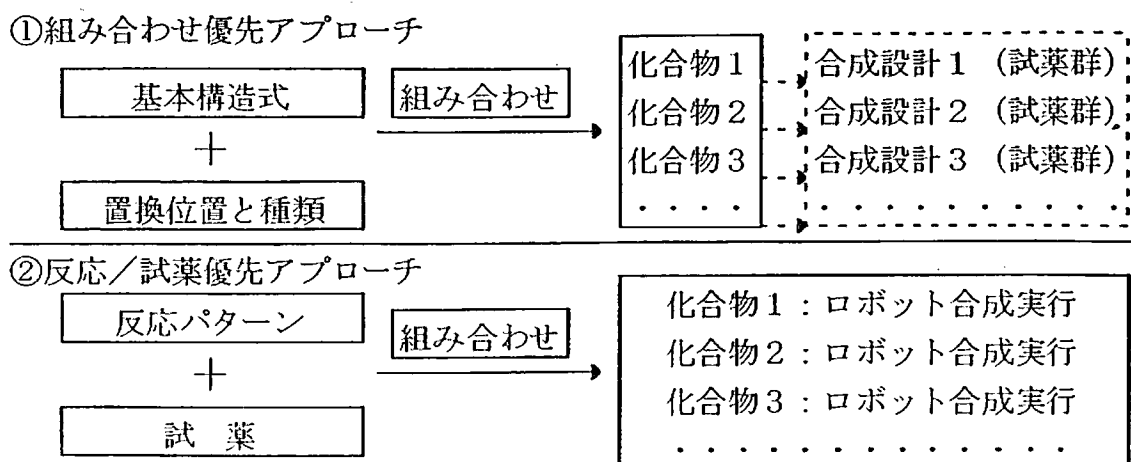


図 . コンビナトリアル合成支援システムの二つのアプローチ

組み合わせ優先のアプローチでは構造式が先にある、合成が後付けとなる。従ってこの場合の合成概念は従来のターゲット指向型の合成と何ら代わる点がない。むしろ合成化合物が多数存在し、複数化合物を同時に、且つ最適に合成するプランを立てる事は従来の合成よりもはるかに困難な作業である。著者は現在合成を行っていないが、化合物構造式を決めるよりも、化合物構造式を決めた後の合成プランにはるかに長い時間がかかるという話は良く耳にする話である。

反応パターンと試薬優先のアプローチは、ターゲット指向というよりは手続き優先の考えに従ったものである。反応で生成される化合物は、反応パターンと試薬を決めた時点で自動的に決まってしまう。しかしこのアプローチであれば研究者に大きな負担を強いる合成ルートのプランニングは必要なく、簡単な手続きだけで合成ロボットを動かす事が可能となる。

ターゲット化合物を厳しく限定する必用がなく、数と早さが勝負となるコンビナトリアルケミストリ/HTS の特性を考えるならば、前節でも述べたように、化合物を合成するのではなく試薬を合成するという考えの方が適しているといえる。この点で考えると、組み合わせ的に総てを網羅する事は出来ないが実際の手続きにおいて大きな負担となる合成プランニングが殆ど必用の無い、合成パターン/試薬優先のアプローチが現実的に有利なアプローチと考える。

試薬会社から入手できず、他社にないユニークな試薬を豊富に揃え、それらを組み合わせで多数の化合物群を高速に合成して HTS にかける。これが著者の提唱する組み合わせ化合物、および組み合わせ試薬の考えである。市販の試薬ライブラリや自社化合物ライブラリを HTS した後は、このような合成パターン/試薬優先アプローチで化合物を合成する事が必用となるであろう。

勿論、組み合わせ的なアプローチによる合成は無くなるわけではない。しかし、単にシステムから創出される網羅的な化合物群を闇雲に合成するのではな

く、構造-活性相関等の技術の適用により厳しく選択/活性予測等行って厳選された化合物だけが、後に合成化学者の手により合成されて HTS にかかる。

7. まとめ

コンビナトリアルケミストリの化合物ライブラリ関連技術とロボットによる合成との繋がり、および組み合わせ試薬の提案についてまとめた。化合物ライブラリの重要さは本文でも述べた通りである。上流にある化合物ライブラリの品質が悪ければ、その汚染の影響は下流にある作業の総てを徒労にする可能性がある。この意味で非常にやりがいのある仕事と考える。また、いくら優れた化合物ライブラリを構築したとしても実際に合成しなければ絵に書いた餅同様、何も生み出してはくれない。合成するにしても、全く新しい発想に立たないと数と早さを勝負とする全く新しい尺度を満たすことは困難である。

現時点におけるアプローチを幾つか紹介したが、著者の考えでは現段階は単に既存の情報化学および構造-活性相関技術をコンビナトリアルケミストリ/HTS に適用したレベルと感じられる。従来の構造-活性相関と同様、まず先に自分なりの解析をしてリード候補化合物ライブラリはこれであると勝手に決めて、後はおまかせ形式のアプローチが多いのがこの根拠である。コンビナトリアルケミストリ/HTS は自動車同様に多種多様の技術の集積体である。一つ一つの部品が勝手に動いては全体の流れに支障をきたす事は明白である。

では、現在のコンビナトリアルケミストリに何が必要なのか？ 著者が見る限り、コンビナトリアルケミストリの実施で今後最大のネックとなるのはやはり合成であると思う。従って、一連の作業はこの合成中心に展開するように設計する事が必要と考える。具体的には、現在 De Novo デザイン等で行っている化合物創出は自分の理論で取り出してきた化合物であり、合成する事を考えたアプローチではない。そのくせ、構造的に変化の大きな化合物群を多数生み出して総てリード候補化合物ですと提案する。これでは実際に合成する研究者はこの結果を素直に受け入れられないであろう。つまり、最初からコンビナトリアルケミストリ/HTS の本質である協調作業の前提が崩壊している事になる。

合成が重要であれば、この合成を上流過程に据えて考える事が必要である。あるいは合成によるフィルターを前提とした構造-活性相関の再設計が必要である。構造-活性相関で勝手に化合物を創出するのではなく、まず最初に合成出来る化合物群は何であるかを議論する。合成する戦略と構造式が決定されたならば、重複合成や無意味な合成等の無駄を省くために構造-活性相関の技術を駆使して最終的な化合物ライブラリ、あるいはリード候補化合物ライブラ

りとする。このような目的に答えるには、従来の構造-活性相関手法には以下のような条件を満たす事が求められる。

- ①構造式の定まった化合物の活性評価/予測機能の充実
- ②合成を前提とした構造-活性相関手法への組み直し
- ③合成支援システムや化合物/反応データベースとの連携強化

この反対に合成の方には単に数合わせや高速合成手法の開発要求のみならず、構造-活性相関との連携を念頭においた時の合成のあり方が求められる。合成の都合で好き勝手に合成するのではなく、幾つかの条件を前提として合成プランニングすることが必要となる。

- ①ファーマコフォア等の使用を優先させて合成設計する
- ②同族体にこだわらず、分子多様性を念頭に置く
- ③化合物のユニット性を高め、より大きな単位での反応を考える

両者の歩み寄り（妥協点？）を見いだすことは容易ではないかも知れないが、コンビナトリアルケミストリ/HTS を育てるにはこれらの問題をクリアすることが必要である。